Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation, Bethesda, Maryland*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

Cognitive Aspects of Information Processing: I. The Familiarity of S-R Sets and Subsets: P. M. FITTS AND G. SWITZER 321
Application of a Stimulus Sampling Model to Children's Concept Formation with and without Overt Correction Responses: P. SUPPES AND R. GINSBERG 330
Magnitude of Reinforcement and Choice Behavior in Children: S. SIEGEL AND J. M. ANDREWS .. 337
Studies in Associative Interference: N. F. BESCH, V. E. THOMPSON, AND A. B. WETZEL 342
Brightness Discriminations with Constant Duration Intermittent Flashes: R. L. ERDMANN 353
The Temporal Course of Proactive Inhibition for Serial Lists: L. POSTMAN 361
Transfer of Training as a Function of Task Difficulty in a Complex Control Situation: D. A. GOLDSTEIN AND J. M. NEWTON 370
Effects of Exposure Time and Density on Visual Symbol Identification: W. H. TEICHNER AND E. SADLER 376
Sensory Integration with and without Reinforcement: V. O. TYLER, JR. 381
A Psychophysical Study of Hunger in the Rat: R. C. BOLLES 387
Keeping Track of Variables That Have Few or Many States: D. B. YNTEMA AND G. E. MUESER 391
Recency and Frequency in Paired-Associate Learning: L. R. PETERSON, D. SALTZMAN, K. HILLNER, AND V. LAND 396
A Test of a General Utility Theory Model for Probability Learning: K. B. LITTLE, Y. BRACKBILL, AND S. H. KASSEL 404
Presentation Time, Total Time, and Mediation in Paired-Associate Learning: B. R. BUGELSKI .. 409
Supplementary Report: Generalization of a Nonverbal Response to Aggressive Verbal Stimuli: J. H. GEER AND A. H. BUSS 413
Supplementary Report: Stimulus Generalization Gradients along a Luminosity Continuum: G. OLSON AND R. A. KING 414
Supplementary Report: Shift from Nonreward to Negatively Correlated Reward: F. A. LOGAN AND L. M. GONZALEZ 416

American Psychological Association

Vol. 63, No. 4 April 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*. When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1961 the publication cost is $30 per printed page.

All **regular** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. **Early publication** articles should be sent to the Editor: Arthur W. Melton; Department of Psychology, University of Michigan; Ann Arbor, Michigan until *July 1, 1962*. Thereafter they should be sent to the Editor-elect. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Acting Production Manager

RUTH P. JAGOLINZER
Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 63, NO. 4 APRIL 1962

COGNITIVE ASPECTS OF INFORMATION PROCESSING: I. THE FAMILIARITY OF S-R SETS AND SUBSETS[1]

PAUL M. FITTS AND GAIL SWITZER

University of Michigan

Choice reaction time (RT) has been shown to be a function of several classes of variables, including discriminability (Crossman, 1955), S-R compatibility (Fitts & Seegar, 1953; Morin & Grant, 1955; Welford, 1958, Ch. 6), and amount of information to be transmitted. The present series of studies was designed to investigate the effects of certain cognitive variables, which can be viewed as influencing Welford's (1960) translational stage of information processing. In particular, the present studies deal with *S*'s ability to adopt and utilize cognitive sets appropriate to the number of alternatives permitted to occur in an experiment. The use of familiar vs. unfamiliar groups of stimuli, *considered as a set*, is emphasized, in contrast to the usual emphasis on the familiarity, or response strengths, of individual S-R associations.

Mental set, expectancy, and related concepts have, of course, been frequently invoked in psychology for many years (see reviews by Dashiell, 1940; Gibson, 1941). For example, Helson (1959) has emphasized the widespread importance of adaptation level as influenced by background and contextual factors, Postman and Bruner (1949) have demonstrated the greater efficiency of perception under single than under multiple sets, and Harlow (1949) has developed the concept of learning sets. However, the importance of cognitive sets in influencing the efficiency of performance in information handling (speeded perceptual motor) tasks has not been systematically investigated.

The importance of cognitive set as a factor influencing information handling efficiency, is indicated by an examination of some apparently conflicting results which have appeared recently from experimental studies of choice RT. One of the most generally accepted empirical laws governing perceptual-motor behavior is the one relating RT to stimulus uncertainty, which has been expressed by some writers as:

$$RT = A + BH_t \qquad [1]$$

H_t is average information transmitted per response.

A and B are constants.

This linear increase in RT with increasing information transmitted, first emphasized by Miller (1951), Hick (1952), and Hyman (1953), has been verified by numerous subsequent experiments. A few studies, however, have revealed instances in which RT is essentially independent of information, i.e., in which the slope constant, B, in Equation 1

[1] This research was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49 (638)-449.

approaches a value of zero. Conditions under which the latter relation has been found appear to be characterized by the use of particular stimulus and response codes, and/or by highly compatible or overlearned S-R sets.

When the response required of *S* is to move one of his fingers, RT usually has been found to increase as H_t increases, over the range of 1 to 10 fingers. However, once *S* is required to use 2 or more fingers in combination, results (Klemmer & Muller, 1953; Seibel, 1960) indicate that variations in the size of the subset of finger combinations used in the task has little or no effect on RT. These effects have been found after small or moderate amounts of training; very extensive training lasting over many weeks may, of course, give different results.

Other instances of very small changes in RT as a function of stimulus uncertainty have been reported when responses were made to English language symbols (words, letters, or numerals). Quastler and Brabb (1956) found very small changes in typing speed when copy was drawn from alphabets of differing size. Similarly, Pierce and Karlin (1957) found that alphabet size per se, over the range from 2 to 256 words, had little effect upon speed of reading lists of random words aloud, provided word length and frequency of general usage in the English language (i.e., familiarity) were held constant. Mowbray (1960) found no difference in RT in naming numbers, over the range from 2 to 10 alternatives. Davis, Moray, and Treisman (1961) played numerals, letters, and words from a tape recording into earphones and required *S*s to speak each sound as soon as possible after hearing it. Responses to numerals were slightly faster (by about 20 msec.) than to the larger alphabet of letters, but no effects of varying the size of the subsets of numerals or of letters were found.

The dependence of the constants in Equation 1 on coding variables is further shown in a study by Brainard, Irby, Fitts, and Alluisi (1962) employing four different S-R ensembles. In three of their four S-R codes a reduction in RT followed a reduction in number of alternatives. In the fourth, however, involving vocal (naming) responses to numerals, no significant changes in RT were found for two, four, or eight alternatives. The authors suggest that this lack of an effect may have been due to the manner in which stimulus uncertainty was reduced. In the two-alternative case, for example, 4 and 7 were used as stimuli and it was noted that this pair does not form a familiar subset. Newman (1959) has offered a similar hypothesis, namely that ". . . short-term instruction is probably not effective in reducing the complexity of the task in the face of a history of many hours of practice . . ." (p. 14).

Leonard (1959), using tactual stimuli applied to the finger tips, and Mowbray and Rhoades (1959), training 1 *S* for several months with lights as stimuli, both found little or no effect of number of alternative finger responses even when responses were made with single fingers. Leonard (1961) hypothesizes that number of alternatives has the least effect when compatibility, discriminability, and/or familiarity are high.

In summary, all of the exceptions to the general finding that RT increases with stimulus uncertainty appear to indicate the importance of coding variables, or, more specifically, the importance of (*a*) the natural or learned relations between individual stimuli and responses, and (*b*) familiarity in dealing with particular sets and subsets of stimuli and responses. The outline of a general theory relating to the latter type of effect can be stated as follows.

First, it appears that facility in dealing with a particular ensemble of stimuli develops slowly and for the most part continuously with practice. Conceptual understanding of the requirements of a task may be rapid, but speed and accuracy of responses will usually continue to improve slowly, over a period often measured in years, depending on the complexity of the situation (alphabet size, laws of syntax, etc.). Thus, speed and accuracy of choice in dealing with a particular event is assumed to depend in part on the amount of previous practice in using the cognitive set appropriate to the over-all task or situation in the context of which the event occurs.

Second, it is assumed that a cognitive set, once it has been thoroughly established, can become dominant in as short a time as a second or less, as a result of the occurrence of an appropriate cue. Since a particular S-R pair may be included in several cognitive sets of differing size, the speed and accuracy of the response to any given stimulus will be in part a function of the set which is dominant at the time the response is made as well as of S-R association strength per se.

Third, it is assumed that, when *S* is asked to perform a new task, some specific or multiple cognitive set will become dominant. This initial cognitive set may include excess or inappropriate stimuli and responses, and some appropriate stimuli and responses may

not be included. In the words of Postman and Bruner (1949) it may not be "finely tuned." Such a cognitive set for a given task can be thought of as analogous to a computer program, governing a data processing task, such as the carrying out of a search procedure or the application of a statistical decision rule, and an associated set of data in a rapid-access memory; an inappropriate cognitive set is then analogous to a program and a set of data not entirely appropriate to the task at hand (i.e., to the type of decision or search required).

The present experiments were planned to test certain of these assumptions, taking advantage of *Ss*' extended experience with sets of English alpha-numeric symbols. Prior to each experimental session *Ss* were always told exactly what stimuli to expect and what responses to make during a session. However, it was postulated that they would be able to prepare adequately for the specific task, in advance, only if they had previously developed a cognitive set appropriate to the specific ensemble of stimuli and responses to be used in the task.

Method

Apparatus

Stimuli were presented by means of a projection-type "in-line" display (Exp. 1) or a double-bladed shutter which exposed white cards carrying black letters (Exp. 2 and 3). The *Ss* sat in front of a large gray screen and stimuli appeared in a center window. They wore Wilson "sound barrier" headsets and listened to broad-band masking noise except when receiving communications from *E*. Onset time of vocal responses was measured by use of a boom microphone, electronic voice key, and .01-sec. time clocks. The lag time of the apparatus (shutter, voice key, relays, etc.) was .04 sec.; this lag was subtracted from *Ss*' response times.

Subjects

The *Ss* were 78 volunteer, university students who served for pay. In Exp. 1 and 3 three groups of 8 *Ss* and three groups of 15 *Ss*, respectively, including men and women, were used. In Exp. 2 one group of 9 *Ss*, all men, was used.

Procedure

Experiment 1.—The 24 *Ss* were pretested on a two-choice color-naming task and mean RT scores were used to form three strata; equal numbers were then assigned randomly from within each stratum to each of the three experimental groups, with the restriction that each group contain the same proportions of men and women.

Group 1-LF used the larger, familiar set of eight numerals, 1 through 8. Group 1-SF used a small familiar subset of two numerals, 1 and 2. Group 1-SU used a small unfamiliar subset of two numerals, 2 and 7 (2 was chosen to provide a stimulus common to all three conditions; 7 was chosen because it had been used by Brainard et al., 1962).

During Sessions 1, 2, and 3 each group was given 80 trials. Stimuli were presented in a different randomized order to each *S*, with the restriction of equal frequencies of alternatives over each session. Stimuli were exposed exactly 2 sec. after a warning signal (a .5-sec. interruption of the noise in the headset) and there was approximately a 10-sec. interval between stimuli. Stimuli disappeared as soon as *S* began to speak. Following the initial instructions *E* communicated with *S* only to give knowledge of results (fastest and slowest times) after each cycle of 10 responses.

Experiment 2.—Each of 9 *Ss* served for 4 days and was tested with each of the following sets of stimuli on each day: (*a*) Numerals 1 and 2; (*b*) Numerals 1, 2, 3, and 4; (*c*) Numerals 4 and 7; (*d*) Numerals 4, 5, 6, and 7; (*e*) Numerals 1 through 8. A total of 200 trials was given during each session, 40 under each of the five conditions. The sequence was randomized and different for each *S* within blocks of 40 trials and counterbalanced across sessions. The *Ss* were told what the set of stimuli would be before each block of trials.

Experiment 3.—Pretests on the color-naming task were given and the 45 *Ss* assigned to three groups by the same method as in Exp. 1.

Group 3-LF used all 26 capital letters of the English alphabet. Group 3-SF used a small familiar subset of three letters, A, B, and C. Group 3-SU used the small unfamiliar subset, E, B, P. The letter B, which was common to all three conditions, was chosen because it is a hard phoneme, well suited for detection by the voice key. The other two letters of the unfamiliar subset, E, and P, were chosen because they have frequencies of usage in written English as well as phoneme frequencies similar to the letters A and C. Discriminability, and interpair confusion

frequencies in legibility studies, are approximately equal (Howell & Kraft, 1959).

Each group was given 80 trials on Sessions 1, 2, and 3. Letter sequences were randomized over sessions, each letter occurring a minimum of three times per session. Other procedures were the same as for Exp. 1.

Results

Reaction times.—Speaking time (time from initiation to completion of the vocal response) was found in preliminary tests to be of short duration and relatively invariant with conditions, when single letters or numerals were used. Therefore only RT data (time to initiate the vocal response) were employed in these experiments. Mean RT scores for each group on each session were computed. All results are consistent with the hypothesized differences between familiar and unfamiliar subsets of stimuli.

The most pertinent data for Exp. 1 are those for the numeral 2, which was common to all three groups (see Fig. 1). On Session 1 the RT to 2, when 2 occurred as an alternate to 7 (Group 1-SU), was as slow as when it occurred as part of a set of eight numerals. But when it occurred as an alternate to 1 (Group 1-SF) the average RT was about .02 sec. less than when it was part of the larger set. Groups 1-SF and 1-SU both showed some learning during Sessions 2 and 3, but the improvement was small. Less learning occurred for Group 1-LF.

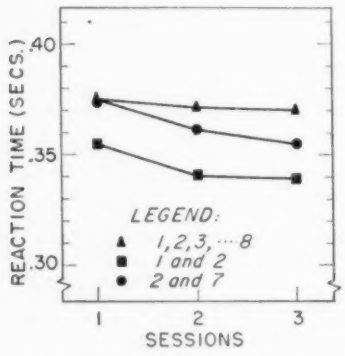

Fig. 1. Reaction times to the numeral 2 when it occurred as a member of three sets of numerals (Exp. 1).

TABLE 1

Analysis of RTs to the Numeral 2 (Exp. 1) When Contained within Different Sets of Stimuli

Source	*df*	*MS*	*F*
Between *Ss*	23		
No. Alternatives (A)	2	47.37	2.49
Error (s)	21	18.96	
Within *Ss*	48		
Sessions (D)	2	0.03	
A × D	4	1.47	2.88*
Error (w)	42	0.51	

* $P < .05$.

Results of an analysis of covariance for these data are given in Table 1 (covariance analysis was used instead of a Treatments × Levels analysis because the stratification employed in assigning *Ss* to groups was quite coarse and did not make full use of the information in the pretests). Color-naming pretest scores correlated highly with test performance for all the three groups ($r = .76$, .74, and .82, respectively). The effect of Conditions was not significant, but the interaction of Sessions and Conditions was significant at the .05 level, indicating differential learning rates.

In Exp. 2, where *Ss* also worked with numerals, but served as their own controls, the results were very similar. Data for familiar sets of stimuli containing the numerals 1 and 2 are summarized in Fig. 2A; results of an analysis of variance for these data are given in Table 2. For this familiar subset condition RT increased slightly, but significantly, for each increase in the number of equiprobable alternatives. In fact

TABLE 2

ANALYSIS OF VARIANCE OF RTs TO NUMERALS WHEN CONTAINED WITHIN SETS OF DIFFERING SIZE AND FAMILIARITY

Source	*df*	Numerals 1 and 2		Numerals 4 and 7	
		MS	*F*	*MS*	*F*
No. Alternatives (A)	2	38.32	12.0**	7.49	2.44
Sessions (D)	3	38.15	11.2**	34.97	5.02**
Subjects (S)	8	115.47		119.28	
A × S	16	3.19		3.07	
D × S	24	3.40		6.97	
A × D	6	2.63	1.43	3.41	1.87
A × D × S	48	1.84		1.82	

** $P < .01$.

8 out of the 9 *S*s gave four-session individual mean RTs which showed a progressive orderly increase as a function of stimulus uncertainty, as Equation 1 would predict. For the unfamiliar subsets, however, RT (for 4 and 7) had no consistent relation to the size of the stimulus set (see Fig. 2B). Not 1 of the 9 *S*s had individual mean RTs for 4 and 7 which increased in an orderly fashion with stimulus uncertainty as predicted by Equation 1. The over-all *F* test for the effect of number of alternatives was not significant (see Table 2). The group means on the last two sessions actually were fastest for the eight-alternative condition and the differences in RT to 1 and 2 vs. 4 and 7, two-choice task, were greater, in fact, at the end than at the beginning of practice. Thus the same *S*s who were consistently able to give faster RTs to a familiar set, as stimulus uncertainty was reduced, were unable to do so when an unfamiliar set was used. The significance of this finding is further indicated (see Fig. 2A and 2B) by the fact that average RTs for these contrasting pairs (1 and 2 vs. 4 and 7) were almost exactly equal when both

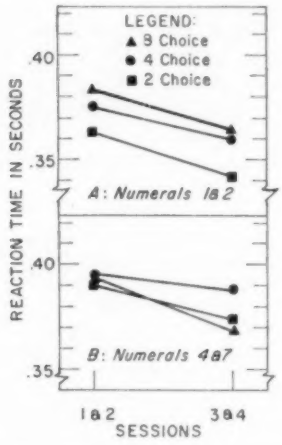

FIG. 2. Reaction times to several sets of numerals (Exp. 2). (In A are shown data for 1 and 2 when used alone and as a part of two other sets; in B are shown similar data for 4 and 7.)

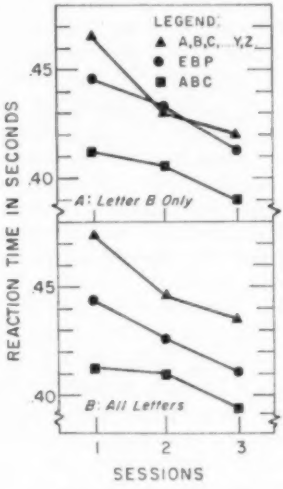

FIG. 3. Reaction times to three sets of letters (Exp. 3). (In A are shown data for the letter B when it occurred in each of these sets; in B are shown the means for all letters comprising the three sets.)

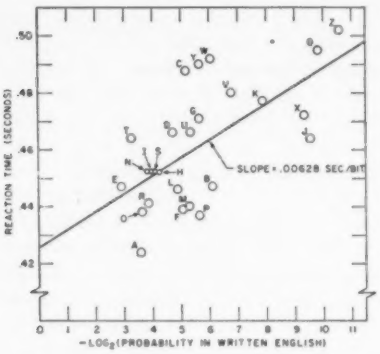

FIG. 4. Relation of English letter frequencies (Pratt, 1939) to vocal RT when all letters were used equally frequently in Exp. 3.

were contained in the same set of eight familiar numerals. Thus the differences were not due to association strength per se.

The results for Exp. 3 (see Fig. 3A), employing letters of the alphabet, are similar to those for Exp. 2. RT scores for saying B were faster when B was one of a 3-letter, familiar subset than when it was part of the entire 26-letter alphabet. When B occurred as one of a small unfamiliar subset, however, mean RT was about the same as when it occurred as one of a large familiar set. An analysis of covariance for these data indicated that the effect of the experimental conditions was significant at the .01 level (see Table 3).

When the scores for *S*s in Exp. 3 were computed across all the letters used by each group (see Fig. 3B), average RTs were slowest for Group 3-LF and next slowest for Group 3-SU. Additional comparisons between the larger and smaller sets of letters are also of interest. RTs to the letters A and C were 40 msec. faster when they occurred in the small familiar set than when they occurred in the large familiar set; however, RTs to E and P were only 3 msec. faster in the small unfamiliar set than in the large familiar set.

Correlations between relative letter frequencies in English (Pratt, 1939) and average RTs, were computed for Group 3-LF, and found to be moderately high (r = .60 for Session 1 and .62 for Session 3). The absolute magnitude of these RT differences, due to English letter frequencies, were of about the same magnitude as the maximum differences produced by reducing the size of the stimulus set in Exp. 3 (e.g., the difference between the most frequent letter (E) and the least frequent one (Z) was about 50 msec.). (See Fig. 4).

Errors.—The over-all percentages of errors in the three experiments were small (1.4%, 0.2%, and 0.6%) and no important differences in proportions of errors occurred under the different experimental conditions. Thus errors were 1.4%, 1.1%, and 1.6% for the three groups in Exp. 1, and 0.4%, 0.9%, and 0.4%, respectively, for the three groups in Exp. 3.

Individual RT distributions.—The distributions of RTs during single sessions were examined for a sample of *S*s. All of the samples appeared to follow essentially a normal distribution (not a log-normal distribution as

TABLE 3

ANALYSIS OF VARIANCE OF RTS TO THE LETTER B (EXP. 3) WHEN CONTAINED WITHIN DIFFERENT SETS OF STIMULI

Source	*df*	*MS*	*F*
Between *S*s	44		
No. Alternatives (A)	2	146.57	5.31**
Error (s)	42	27.61	
Within *S*s	90		
Sessions (D)	2	1.23	—
A × D	4	7.01	—
Error (w)	84	6.35	

** $P < .01$.

might have been expected). Practically all the mean RTs of all *S*s fell within a range of ±100 msec.

Discussion

The results substantiate the prediction that a stimulus element, in a familiar subset, will elicit faster RTs than will the same element either when it occurs as part of a less familiar stimulus subset of the same size or as part of a large and equally familiar set. Demonstration of such effects requires careful control of differences in discriminability and association strengths of stimuli and responses; these are believed to have been adequate in the present experiments. Fully as important an aspect of the present findings, however, is the relatively small absolute magnitude of the obtained differences.

In some tasks relatively large increases in choice times result from increases in stimulus uncertainty. Hyman (1953), for example, using lights as stimuli for vocal responses and employing 4 highly practiced *S*s, found a median increase in RT of about .176 sec. per bit of stimulus uncertainty. Brainard et al. (1962), using groups of 12 naive *S*s, found an increase in RT of .178 sec. per bit for a similar S-R code (vocal-numeral responses to lights). Hick (1952), for 1 highly practiced *S*, found an increase of about .11 sec. per bit for finger responses to lights (in the range up to 10 alternatives), whereas Merkel's data, reported by Hick, had a slope of about .13 sec. per bit. Brainard et al. (1962) found an average increase of .136 sec. per bit for their naive *S*s when an S-R code similar to Hick's was used. In the present study, however, the maximum difference in RT for two vs. eight numerals was only .03 sec., or .015 sec. per bit of stimulus uncertainty. Differences resulting from the use of 3 vs. 26 letters were also very small. Thus the present results are similar to those reported by investigators who have studied responses to English language symbols. Even when we compare numerals with letters, two highly familiar alphabets of differing size, the difference in average vocal RT is not great—.092 sec. on Session 1 and .062 on Session 3 for all stimuli in each set. The increase in RTs in going from numerals to letters is thus within the range from .054 to .036 sec. per bit, which is greatly below the values reported when less familiar coding alphabets are used.

The generalization appears warranted, therefore, that when *S* is asked to give stereotyped responses to highly familiar sets of symbols, such as those used in English, his information-handling efficiency will already be very high, and his RT will be reduced only slightly by temporary reductions in alphabet size, even when the temporary reduction in uncertainty makes maximum use of familiar subsets.

The present results form the basis for a better understanding of the role of learning in choice behavior. All recent attempts at development of a theoretical model to account for the increase of RT with increases in stimulus uncertainty assume the efficient or selective use of memory (see Welford, 1960), but none deal specifically with the nature of the selectivity of the memory process. This is true both of models which propose that a simultaneous or successive search process (Hick, 1952; Rappaport, 1959) or that a sequential statistical decision process (Stone, 1960; Tanner, 1961) is responsible for the relation between increased time and increased number of alternatives. Thus, Rappaport (1959) assumes that "the task is to match the stimulus perceived with one of the n stimuli stored in the memory" (p. 301), and Stone (1961) assumes that *S* is familiar with the probability distributions of the alternative events and that "such familiarity may be the result of a process of learning, provided *S* has performed many trials" (p. 255). The cognitive set theory proposed here is an attempt to specify more explicitly some of the characteristics of the process that makes a selective portion of *S*'s memory available in a particular choice situation. In terms of a computer analogy, the establishment of an appropriate cognitive

set for a particular task is like transferring a portion of the total stored information relating to previous signals to a rapid-access memory, and providing appropriate instructions for effecting a given (search or statistical decision) operation which will require comparison of incoming signals with the information stored in the rapid-access memory.

There remains the question of why in certain experiments, particularly those of Leonard (1959), Mowbray and Rhoades (1959), and Mowbray (1960), RT shows no change at all as the size of the set of alternatives is changed. In all three of these experiments responses were either highly familiar and/or compatible ones. It is possible that a different type of choice mechanism becomes operative in such cases. For example, a search or statistical decision mechanism might be replaced by a direct association mechanism that is independent of the number of alternatives. However, it should be noted that in the first two of these studies the same *S*s served under all conditions; thus another possible explanation is that *S*s may have developed an undifferentiated cognitive set across all the experimental conditions, and not changed this cognitive set when *E* changed conditions.

The results from the recent study by Mowbray (1960) are particularly relevant to the present discussion. He trained 150 *S*s, in five groups. Each group served under a single condition (2, 4, 6, 8, or 10 alternatives) for one session and each *S* made a total of 60 responses to each stimulus in his set. Stimuli were numerals and the response was to vocalize the familiar name. All groups, regardless of number of alternatives, gave a mean RT of about .47 sec., and Mowbray (1960) concluded that ". . . the results support the hypothesis that choice reaction times for practiced responses are constant" (p. 200) over the range from 2 to 10 alternatives. The significant fact, however, is that all of the subsets of numerals which Mowbray[2] used were relatively unfamiliar ones: 4, 8 for two-choice; 1, 4, 6, 8, for four-choice; 1, 2, 3, 5, 7, 8, for six-choice; etc. Thus Mowbray's results are actually in full agreement with those obtained in the present experiments and are consistent with the predictions made from a cognitive set theory.

[2] G. H. Mowbray, personal communication.

Summary

The investigation concerned *S*'s ability to reduce his reaction time (RT) on the basis of knowledge that only a small subset of a larger, familiar alphabet would occur as stimuli in a particular experiment. It was predicted that *S*s could make effective use of this knowledge only when the smaller subset itself constituted a highly familiar group. English letters and numerals were used as stimuli and familiar as well as unfamiliar subsets were employed; the response was to vocalize the familiar name of the stimulus.

The results of three experiments were in agreement with the predictions. RTs were faster for small, familiar subsets (such as 1, 2; A, B, C); RTs for small, unfamiliar subsets (such as 2, 7; 4, 7; E, B, P) were about as slow as for the larger familiar sets of which they normally were a part. In no instance, however, was the absolute reduction in RT resulting from a decrease in number of alternatives very large for English language alphabets. It therefore appears that large differences in RT may result from changes in number of alternatives only when tasks are low in S-R compatibility, such as tasks using novel S-R codes or requiring novel information transformations.

These results support the view that information-handling rate is in part a function of cognitive sets which reflect the preparation which *S* makes, in advance, for responding to any one of a group of stimuli. Some characteristics of these cognitive sets are discussed, and the notion is related to recent theories of choice behavior.

References

Brainard, R. W., Irby, T. S., Fitts, P. M., & Alluisi, E. A. Some variables influencing the rate of gain of information. *J. exp. Psychol.*, 1962, **63**, 105–110.

Crossman, E. R. F. W. The measurement of discriminability. *Quart. J. exp. Psychol.*, 1955, **7**, 176–195.

Dashiell, J. F. A neglected fourth dimension to psychological research. *Psych. Rev.*, 1940, **47**, 289–305.

DAVIS, R., MORAY, N., & TREISMAN, A. Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, 1961, **13**, 78–89.

FITTS, P. M., & SEEGAR, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *J. exp. Psychol.*, 1953, **46**, 199–210.

GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, **38**, 781–817.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HELSON, H. W. Adaptation level theory. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959.

HICK, W. E. On the rate of gain of information. *Quart. J. exp. Psychol.*, 1952, **4**, 11–26.

HOWELL, W. C., & KRAFT, C. L. Size, blur, and contrast as variables affecting the legibility of alpha-numeric symbols. *USAF WADC tech. Rep.*, 1959, No. TR 59-536.

HYMAN, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.

KLEMMER, E. T., & MULLER, P. F. The rate of handling information. Key pressing responses to light patterns. *USAF HFORL memo. Rep.* 1953, No. 34.

LEONARD, J. A. Tactual choice reactions. *Quart. J. exp. Psychol.*, 1959, **11**, 76–83.

LEONARD, J. A. Choice reaction time experiments and information theory. *Proc. 4th Inform. Theory Conf.*, 1961.

MILLER, G. A. *Language and communication.* New York: McGraw-Hill, 1951.

MORIN, R. E., & GRANT, D. A. Learning and performance of a key-pressing task as a function of the degree of spatial stimulus-response correspondence. *J. exp. Psychol.*, 1955, **49**, 39–47.

MOWBRAY, G. H. Choice reaction time for skilled responses. *Quart. J. exp. Psychol.*, 1960, **12**, 193–202.

MOWBRAY, G. H., & RHOADES, M. U. On the reduction of choice-reaction times with practice. *Quart. J. exp. Psychol.*, 1959, **11**, 16–23.

NEWMAN, E. B. Men and information: A psychologist's view. *Nuovo Cim.*, 1959, **13**, 539–559.

POSTMAN, L., & BRUNER, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.

PRATT, F. *Secret and urgent.* New York: Bobbs-Merrill, 1939.

PIERCE, J. R., & KARLIN, J. E. Reading rates and the information rate of a human channel. *Bell Sys. tech. J.*, 1957, **36**, 497–516.

QUASTLER, H., & BRABB, B. Human performance in information transmission. V. The force of habit. *U. Ill. Control Sys. Lab. Rep.*, 1956, No. R-70.

RAPPAPORT, A. A study of disjunctive reaction times. *Behav. Sci.*, 1959, **4**, 299–315.

SEIBEL, R. Some factors influencing choice reaction time. Paper read at the American Psychological Association meetings, Chicago, 1960.

STONE, N. Models of choice-reaction time. *Psychometrika*, 1960, **25**, 251–260.

TANNER, W. P. Physiological implications of psychophysical data. *Ann. N. Y. Acad. Sci.*, 1961, **89**, 752–765.

WELFORD, A. T. *Ageing and human skill.* Oxford: Oxford Univer. Press, 1958.

WELFORD, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, 1960, **3**, 189–229.

(Received January 14, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 330–336

APPLICATION OF A STIMULUS SAMPLING MODEL TO CHILDREN'S CONCEPT FORMATION WITH AND WITHOUT OVERT CORRECTION RESPONSES[1]

PATRICK SUPPES AND ROSE GINSBERG

Institute for Mathematical Studies in the Social Sciences, Stanford University

In the present experiment, the first of a series of experiments intended to study concept formation in young children, 5- and 6-yr-old *S*s are required to learn two fairly difficult concepts, each represented by three stimuli. The concepts to be learned are numbers in binary notation and the stimuli representing them are varied by the use of different symbols for the binary digits. A feature of the experimental situation is that it permits the application of a simple sampling model to the data and in addition suggests two methods of analysis, one in terms of paired-associate (PA) learning and the other, which we shall later refer to as "pure property analysis," more directly concerned with acquisition of the concepts involved.

If, in analyzing the present data the six stimuli representing two concepts are treated as if they were independent, the situation is analogous to a PA learning experiment. The data then become amenable to analysis in terms of a statistical learning theory that has been successfully applied to PA experiments (Bower, 1961; Estes, 1961).

In the second analysis the three stimuli representing the same concept are treated as if they were identical—in effect one stimulus. The learning situation is then reduced to the association of two stimuli and the two appropriate responses and the PA model used in the first analysis may again be applied to the rearranged data. In this second method of analysis the learning situation is viewed as involving the association between the common property of the stimuli which determines a concept and the identification of that concept, which is the correct response. To distinguish the two approaches the first method—six independent stimuli and two responses—is designated the "paired-associate" analysis, and the second approach—two independent stimuli and two responses—the "pure property" analysis.

A further aim in the present experiment is to establish the role in young children's learning behavior of an overt correction response following an incorrect response. In two-response situations with adult human *S*s, when *S* is informed whether he is correct or incorrect on each trial, the probability of making a correct response is affected by the outcome of a trial rather than the specific response made on that trial. Requiring *S* to make an overt correction response after reinforcement neither improves his learning rate nor influences his asymptotic behavior (Burke, Estes, & Hellyer, 1954). The present experiment attempted to establish whether an analogous situation obtains when *S*s are young children.

[1] The research reported was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare. We are indebted to W. K. Estes for a number of constructive suggestions.

METHOD

Subjects.—There were 48 *Ss* divided between two groups. Each group consisted of 12 *Ss* from Grade 1 and 12 *Ss* from the kindergarten class of Stanford Elementary school. The average age of the former was 6 yr. and 3 mo., of the latter 5 yr. and 1 mo.

Design.—On each trial a stimulus was presented to which *S* made one of two possible responses. The *E* always indicated whether *S*'s choice on that trial was correct or incorrect. For 24 *Ss* (Group C) a wrong choice was followed by an overt correction response; for the remaining 24 *Ss* (Group NC) no overt correction response was required.

Equipment.—The stimuli, printed symbols on $8\frac{1}{2} \times 11$ in. white cards, were presented one at a time on a small metal reading stand to the front of which a sizeable white hook was attached. Two responses were available, the Arabic numerals 4 and 5 each printed in $\frac{1}{4}$-in. red tape on a $2\frac{1}{2} \times 4$ in. white card. A 1-in. composition ring attached to the top of each response card enabled *S* to hang the chosen response on the reading stand hook.

Six different stimuli were presented. Three of the stimuli represented 4 in the binary number system, the remaining three represented 5. The equivalent binary numbers are, of course, 100 and 101, but as quite small children are often familiar with some combinations of 1 and 0, different symbols were used in place of the Arabic numerals. The stimuli used and their Arabic numeral equivalents are shown in Table 1. Sixteen copies of each stimulus were prepared, making 96 stimuli in all. Four different random sequences were used and matched across groups.

Procedure.—The *Ss* were run individually. Each *S* was seated on a small chair with the reading stand on a low table in front of him. A response card was placed in front of and to the right and left of the reading stand (the position, left or right, of the responses was balanced across *Ss*). The *Ss* were told that the numbers on the response cards were 4 or 5. All the children were familiar with these Arabic numerals although recognition was not, of course, necessary to complete the task. It was explained that all the "pictures" to be shown were different ways of writing 4 or 5; that is, they were all the "same as" 4 or the "same as" 5. The *Ss* were instructed to guess whether a picture presented on the reading stand was the same as 4 or 5 and to indicate their guesses by putting the chosen response on the hook in front of the picture.

A practice card was then placed on the reading stand; 2 sec. later *E* said "now" and *S* placed the chosen response card on the hook. The *E* said "right" if the answer was correct and "no" if the answer was incorrect. In Group NC (noncorrection) this completed a trial; in Group C (correction), after an incorrect response, the response card was taken from the hook and *S* was given time to place the correct card on the hook. The trial was terminated by *E* saying "right." In both groups at the end of a trial *S* was allowed to remove the response card from the hook while *E* changed stimuli. The correct intertrial position for the response cards was established before the next stimulus was exposed. A fairly constant speed for each movement on each trial was in general easily established both for *S* and *E*. For all *Ss* it was not necessary after the first few trials for *E* to initiate *S*'s response by saying "now."

TABLE 1

STIMULI AND EQUIVALENT ARABIC NUMERALS USED IN LEARNING AND TEST TRIALS

Trials	Stimuli	Numerals
Learning	$\Gamma^{**}\ \ \lambda\pi\pi\ \ \leftrightarrow\Sigma\Sigma$	4
	$\Gamma^{*}\Gamma\ \ \lambda\pi\lambda\ \ \leftrightarrow\Sigma\leftrightarrow$	5
Test	$\delta\epsilon\epsilon\ \ \alpha\Delta\Delta$	4
	$\delta\epsilon\delta\ \ \alpha\Delta\alpha$	5

To demonstrate the procedure to *S*, two practice trials were given. Each of the stimuli used on these trials consisted of two different objects (dissimilar also to the symbols used in the learning trials) and the response to the first stimulus was always judged as correct, to the second stimulus as incorrect. The order of presentation of practice stimuli was reversed for half the *Ss* across both groups and grades.

The 96 training trials following the practice trials were broken up by a 2-min. rest period between the first and second 48 trials. The learning trials were succeeded by a second 2-min. rest period after which 4 test trials were given. The test stimuli, two of which represented 4 and two represented 5, are shown in Table 1 together with the correct responses. No information of any kind was given on these four trials but at the end of this short session all *Ss* were told they had answered correctly on all trials.

RESULTS AND DISCUSSION

Mean learning curves.—As there was no significant difference between the percentages of correct responses for

kindergarten and Grade 1 *Ss* (the total numbers of correct responses for the two grades being, respectively, 1526 and 1534), the data from both grades were combined and are presented in terms of proportion of errors for Groups C and NC in Fig. 1. The t of 4.00 ($df = 46$) computed between overall responses for the two groups is significant at the .001 level. Although Group C does significantly better than Group NC, *Ss* in the former group do not achieve perfect learning. As a matter of interest both groups were divided into *Ss* who reached a criterion of at least five correct responses on each of the last three presentations of the six stimuli, and *Ss* who did not achieve that criterion. In Group C, 6 of the Grade 1 *Ss* and 6 of the kindergarten *Ss* reached criterion; in Group NC, 2 Grade 1 and 2 kindergarten *Ss* reached criterion. Figure 2 describes the correct responses of criterion and noncriterion *Ss* in Groups C and NC. The difference in learning rate between *Ss* not reaching criterion in Groups C and NC is apparently considerably affected by the correction response variable. A t of 2.75 ($df = 30$) between the overall correct responses of these two subgroups was significant at the .01 level. These results, both for the full groups and the subgroups, indicate that the making of an overt correction response has a positive additional reinforcing effect in young children. This is contrary to the Burke et al. (1954) results for adults.

Theory.—Bower (1961) has recently presented a one-element model for PA learning from which he derived an extensive number of predictions. (A mathematically equivalent model was first formulated by Bush and Mosteller, 1959; see also Estes, 1961.) In effect, it is assumed that in a PA experiment each item may be represented by one stimulus element which may be in one of two states, conditioned to the correct response or unconditioned. On each trial the stimulus associated with the

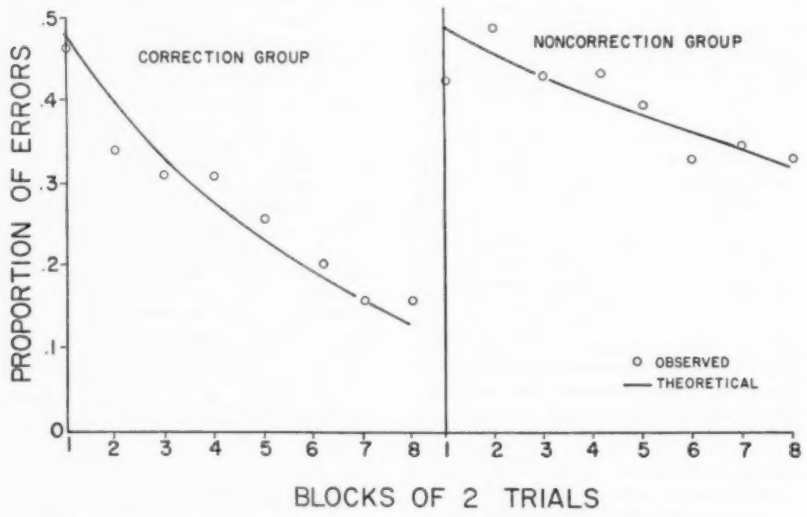

FIG. 1. Observed errors for the 16 trials on each item, averaged over *Ss* and items, and predicted curves from the paired-associate analysis.

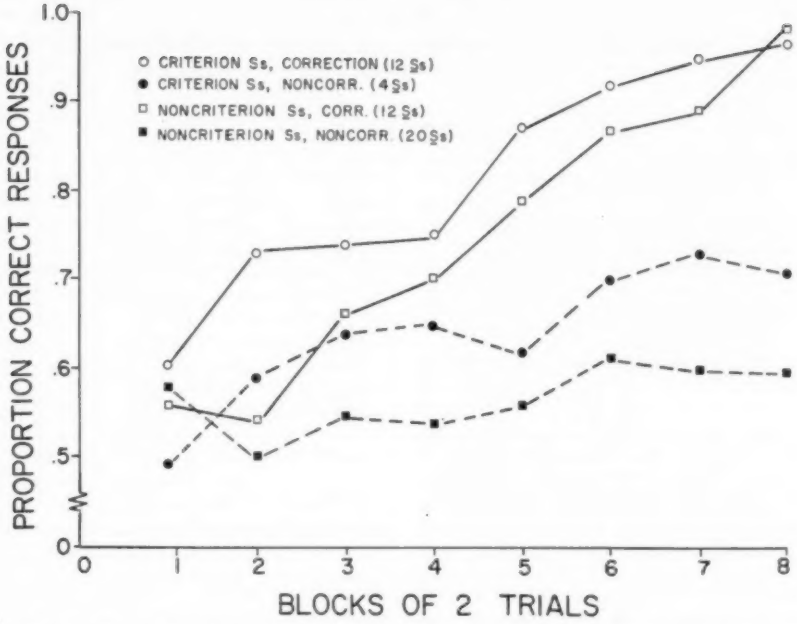

FIG. 2. Proportion of correct responses for *Ss* with at least five out of six correct on each of the last three trials (criterion *Ss*) and for "noncriterion" *Ss*.

item presented is sampled. If the stimulus is unconditioned and the trial is reinforced, it becomes conditioned to the correct response with probability θ and remains unconditioned with probability $1 - \theta$. At the outset of the experiment the stimulus is assumed to be unconditioned. The probability of a response if an unconditioned stimulus is sampled is $1/k$ where k is the number of available responses. The probability of a correct response if a conditioned stimulus is sampled is unity. The reader will recognize that this model is a special case of stimulus sampling theory when the presence of exactly one stimulus element is assumed.

The statistics which Bower derives are presented in some detail and are not therefore developed here. However, since many of the derived expressions assume a larger number of trials or a larger value of θ, they are not appropriate in the present case where both n and θ are small. Consequently wherever necessary we have re-estimated the statistics for finite n and small θ. Some but not all of the formulas we used are given in Estes (1961).

Paired-associate analysis.—Using the one-element model described, with each stimulus treated as an independent item, the single parameter θ was estimated from the number of errors. The value of θ for Group C was .088, for Group NC, .031. The theoretical mean error curves are presented in Fig. 1.

Table 2 presents some predicted and observed statistics for Groups C and NC. For both groups the model, judged from the statistics presented, describes the data rather well. The sums of squared deviations between predicted and observed statistics are nearly the same for both groups, being .118 for Group C and .122 for Group NC.

Certain observed and predicted joint probabilities are given in Table 3. Designating the probability of response i on trial m and response j on trial $m + 1$

TABLE 2

PREDICTED AND OBSERVED STATISTICS FOR GROUPS C AND NC: PA ANALYSIS

Statistic	Group C		Group NC	
	Pred.	Obs.	Pred.	Obs.
Expected errors per item per S		4.32		6.37
SD of errors per item per S	3.13	3.00	3.05	2.97
Expected errors before first success	.92	.94	.97	.89
SD of errors before first success	1.30	1.29	1.36	1.44
Expected number of error runs	2.44	2.48	3.42	3.53
Expected error runs of Length 1	1.36	1.48	1.85	2.00
Expected error runs of Length 2	.59	.52	.86	.78
Expected error runs of Length 3	.27	.20	.40	.41
Expected number of alternations of success and failure	4.25	4.31	6.07	6.20
Expected errors on Trials n and $n+k$ jointly				
(a) $k=1$	1.94	1.86	2.94	2.87
(b) $k=2$	1.71	1.51	2.69	2.67
Net difference between success following error and error following error	.375	.486	.188	.306
Net difference between error following error and error following success	.000	.111	.000	.160
Conditional probability of success following error	.544	.558	.516	.525

as $P_m(ij)$ for $i,j = 1,2$ (where "1" denotes the correct response, "2" the incorrect response), and letting:

$$\bar{P}(ij) = \frac{1}{n-1}\sum_{m=1}^{n-1} P_m(ij)$$

it can be shown that:

$$P_m(11) = 1-(1-\theta)^{m-1}(3/4-\theta/4) \quad [1]$$

and:

$$\bar{P}(11) = 1/(n-1) \quad [2]$$
$$\times \sum_{m=1}^{n-1}[1-(1-\theta)^{m-1} \times (3/4-\theta/4)]$$
$$= 1-(3/4-\theta/4) \times \left[\frac{1-(1-\theta)^{n-1}}{(n-1)\theta}\right]$$

and similarly:

$$\bar{P}(12) = \bar{P}(22) = (1-\theta) \quad [3]$$
$$\times \left[\frac{1-(1-\theta)^{n-1}}{4(n-1)\theta}\right]$$

$$\bar{P}(21) = (1+\theta)\left[\frac{1-(1-\theta)^{n-1}}{4(n-1)\theta}\right] \quad [4]$$

The $\bar{P}(ij)$ values, calculated with the value of θ obtained from the mean error curve, are presented in Table 3 together with the observed $\bar{P}(ij)$. Again the fit for both groups is good. In addition, θ was estimated directly from each of the four observed joint probabilities for Group C. The results were $\hat{\theta}_{11} = .089$, $\hat{\theta}_{12} = .086$, $\hat{\theta}_{21} = .086$, and $\hat{\theta}_{22} = .094$. The mean of these four values is .089, which is surprisingly close to the mean error curve estimate of .088. The closeness of these two estimates is encouraging because in many probabilistic learning experiments the value of θ estimated

TABLE 3

AVERAGE JOINT RESPONSE PROBABILITIES $P(ij)$

$P(ij)$	Group C		Group NC	
	Predicted	Observed	Predicted	Observed
$\bar{P}(11)$	.587	.589	.399	.394
$\bar{P}(12)$	.129	.131	.196	.202
$\bar{P}(21)$	.155	.156	.209	.212
$\bar{P}(22)$	.129	.124	.196	.192

from the mean learning curve is smaller by a factor of from 3 to 10 than the value estimated from sequential data.

Pure property analysis.—The one-element model described was again applied to the data, but in this instance it was assumed that the stimuli representing a single concept could be treated as a single stimulus. For example, there were three items to which the correct response was the Arabic numeral 4. These three items were treated as if they were identical, representing one stimulus to which *S* must learn to associate the response 4. Therefore, instead of six items as in the previous analysis, there were now two items—each to be associated with one of the two responses available. The one-element model as described above was used to analyze the data in this form for Group C only, for reasons to appear later.

The single parameter θ estimated from the number of errors was .028 for Group C. As the present analysis is concerned with two distinct items, as compared with the six items of the PA analysis, it follows that the method of estimating θ from the mean errors will give a parameter value approximately one-third that of the estimate for the PA approach. The theoretical and empirical curves for Group C, plotted over blocks of four trials, are presented in Fig. 3.

Table 4 lists some of the obtained and predicted statistics for the data analyzed for pure property learning. The approximate fit of the model is not bad, but it does not compare with the fit of the PA model.

In view of the excellent fit of the PA model to the present data and the comparative lack of success of the pure property model, the immediate question is whether the experimental situation, in fact, involved concept formation learning, or whether it was reduced by the difficulty of the task to a PA situation. If *S*s were learning a concept rather than, or in addition to, association of specific stimulus and response, the probability of a correct response when *S* is faced with a completely new stimulus illustrating the same concept should be at least greater than chance. For both Groups C and NC the initial probability of a correct response on the first experimental trial of the original learning session was .5. Over the 4 test trials introduced after the 96 learning trials, the propor-

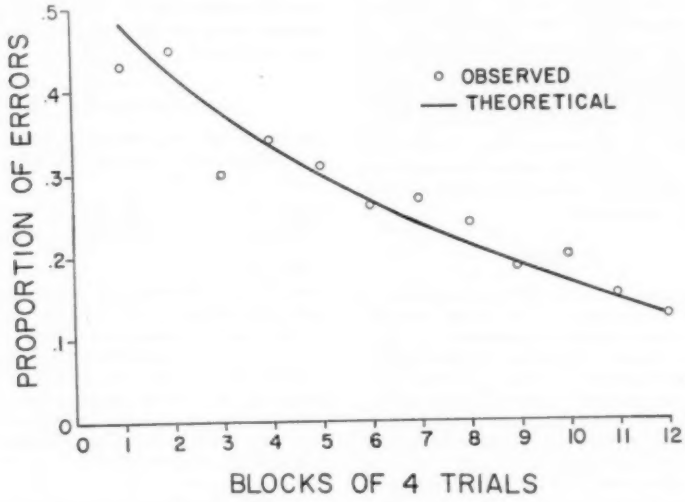

FIG. 3. Observed errors and predicted curve for Group C from pure property analysis.

TABLE 4

PREDICTED AND OBSERVED STATISTICS FOR GROUP C: PURE PROPERTY ANALYSIS

Statistic	Predicted	Observed
Expected errors per item per *S*		13.12
SD of errors per item per *S*	8.85	6.29
Expected errors before first success	.972	1.06
SD of errors before first success	1.36	1.49
Expected number of error runs	6.78	8.50
Error runs of Length 1	5.76	5.57
Expected number of alternations, successes, and failures	13.06	16.25
Expected errors on Trials n and $n + 1$ jointly	6.34	4.48

tion of correct responses for Group C was .646, for Group NC .479. The obtained value for Group C is significantly different from chance at the .01 level ($t = 2.76$, $df = 23$). The fact that the proportion of correct responses for Group NC on the test trials was at the level of chance is the reason the pure property analysis was not applied to this group.

The better fit of the PA model in comparison with the pure property model for Group C and the somewhat paradoxical evidence from test trials that learning had occurred with respect to the property do not seem inconsistent to us, but rather appear to have a natural psychological explanation. Initially *S*s were mastering the problem by acquiring the six individual S-R associations. At a slower rate, because of its greater difficulty, at least some *S*s were also mastering the problem of learning the two general properties, to which they responded on the test trials. Additional experiments will be needed to study in more detail the interaction of these two learning processes.

SUMMARY

Five- and 6-yr-old *S*s were required to learn two concepts in the binary number system, each concept represented by three different stimuli. In Group C (correction) each *S* was required to make an overt correction response after an incorrect response, in Group NC (noncorrection) *S* was simply informed at the end of a trial whether he had responded correctly or incorrectly. The results indicate that Group C performed significantly better than Group NC.

Two methods of analysis were used. For the paired-associate (PA) analysis the stimuli were treated as if they were independent items and for the pure property analysis, all stimuli describing a single concept were treated as if they were identical items. A simple sampling model for PA learning was applied to the data in each case; a good fit of the model to the data was obtained for the PA analysis, the pure property analysis was performed only with the data of Group C and did not approach the adequacy of the PA analysis.

Test trials on new stimuli following the learning trials provided evidence that some *S*s in Group C were learning the concept involved. It was pointed out that evidence of concept learning is not incompatible with success of a PA analysis in the present experimental situation because the learning of the concept may occur after S-R associations have been established.

REFERENCES

BOWER, G. H. Application of a model to paired associate learning. *Psychometrika*, 1961, **26**, 255–280.

BURKE, C. J., ESTES, W. K., & HELLYER, S. Rate of verbal conditioning in relation to stimulus variability. *J. exp. Psychol.*, 1954, **48**, 153–161.

BUSH, R. R., & MOSTELLER, F. A comparison of eight models. In R. R. Bush and W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford, Calif.: Stanford Univer. Press, 1959. Pp. 293–307.

ESTES, W. K. New developments in statistical behavior theory: Differential tests of axioms for associative learning. *Psychometrika*, 1961, **26**, 73–84.

(Received December 14, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 337–341

MAGNITUDE OF REINFORCEMENT AND CHOICE BEHAVIOR IN CHILDREN [1]

SIDNEY SIEGEL AND JULIA McMICHAEL ANDREWS

Pennsylvania State University

In the classical two-choice uncertain outcome experimental situation, first used by Humphreys (1939), *S* predicts, for a series of trials, which of two events will occur on each trial. The events occur with fixed, unequal probabilities, π and $(1 - \pi)$, in a random sequence, the structure of which is in no way contingent upon *S*'s behavior. By witnessing the events as they occur, *S* sees for himself the accuracy of his predictions.

Both Estes' statistical learning model (Estes & Straughan, 1954) and a decision-making model developed by Siegel (1959) yield predictions of asymptotic choice behavior in this situation. The Estes model predicts that, after learning takes place, *S*s tend to predict the two events in the proportion of their actual occurrence. Siegel's Model I, based on considerations from decision-making theory, provides for other modes of asymptotic response as well as for the "matching" behavior. This model yields the prediction that the behavior of *S*s in the two-choice situation is a function not only of the event probabilities but also of the *amount of reinforcement*[2] present in the situation.

Basic to the Siegel model is the hypothesis of *maximization of expected utility*: a person behaves in a choice situation as if he were attempting to maximize the expected subjective value of the outcome of his behavior. Particularly important, on the basis of this hypothesis, are the two concepts of the utility (subjective value) of a correct response and the utility of variability, or the negative utility of "boredom" (Siegel, 1959, 1961).

The prediction of asymptotic response probability, p, yielded by the model[3] may be expressed as follows (Siegel, 1959, p. 307):

$$p = \alpha(\pi - \tfrac{1}{2}) + \tfrac{1}{2}$$

π = the probability of occurrence of the more frequent event.

p = the proportion of times the subject chooses the more frequent event;

and

$$\alpha = \frac{a}{b}$$

a = the marginal utility of a correct prediction.

b = the marginal utility of variability.

As can be seen, the model provides for a change in *S*'s mode of response as an increase in amount of reinforcement for a correct prediction produces an increase in the marginal utility of such a prediction for *S*. It is also seen that in the case when the marginal utility of a correct prediction, a, is equal to the marginal utility of variability, b, the Siegel model predicts the same asymptotic behavior as does the Estes model. Such a case would seem to obtain when knowledge of the outcome of his predictions is *S*'s only source of reinforcement.

Findings in support of both the Estes model and the Siegel model have been reported (e.g., Estes & Straughan, 1954; Siegel, 1961; Siegel & Goldstein, 1959).

[1] Support for this study was provided in part by grants to S. Siegel from the National Science Foundation (G-7071) and from the Central Fund for Research of the Pennsylvania State University, and in part by a Public Health Service Research Fellowship to J. M. Andrews from the National Institute of Mental Health.

[2] It will be noted that the term "amount of reinforcement" is used here with reference to the particular reinforcement for a correct response rather than to the probability of reinforcement, π.

[3] A generalization of the Siegel model for k alternatives has been developed (Siegel & McMichael, 1960). The above expression of the prediction of the model may be derived as a special case of the general model, the case for which $k = 2$, with equal amounts of reinforcement for the two alternatives.

The behavior of children in probabilistic situations has been a specific concern of both choice-behavior investigators and developmental psychologists (e.g., Atkinson, Sommer, & Sterman, 1960; Messick & Solley, 1957; Stevenson & Weir, 1959). Among the research orientations of these investigators, as among those of investigators of adult choice behavior, there has been considerable variation in the degree to which amount of reinforcement is considered a relevant or significant variable in the research setting. It is suggested here that a portion of the variability observed both within and between investigations of children's choice behavior may be a function of this variable, amount of reinforcement.

For the present study preschool children have been observed under two conditions of amount of reinforcement in a suitable modification of the two-choice uncertain outcome paradigm in order to determine the applicability of Siegel's model for choice behavior in this situation to the behavior of young children.

It is hypothesized that the behavior is in part a function of the amount of reinforcement provided for a correct response, as well as of the event probabilities. The prediction is that an increase in the amount of reinforcement for a correct prediction, by producing an increase in a child's marginal utility of a correct prediction, leads to an increased tendency for a child to maximize the probability of such a prediction.

METHOD

Subjects.—The *S*s were 24 boys ranging in age from 3–10 to 5–0, a sample including both nursery-school and nonnursery-school children.

Experimental conditions.—The two-choice uncertain outcome situation was presented to each child in the form of a game in which he predicted, for a series of trials, which of two identical containers placed before him contained an object. After each prediction the contents or lack of contents of the chosen container was revealed to the child. The container in one of the positions held the object more frequently than that in the other position.

Two conditions of amount of reinforcement were used. In the low reward condition, the object used in the series was a button, selected to have minimum interest value for the children. The same button was used throughout the series. Knowledge of the outcome of each trial, as verified by *S*'s indication at the end of the trial as to which container had held the object, was the only experimental reinforcement. Use of verbal reinforcement was avoided by *E*.

In the high reward condition, the object was a small prize which *S* could have to keep if he predicted correctly which container held it. The prizes for the entire series included a wide variety of toys, candy, plastic trinkets, crayons, balloons, coins, and other similar items, rewards of the type shown to have high reinforcement value for nursery-school-age children (Bijou & Sturges, 1959). For each trial, then, there was for *S* the possibility of receiving a prize, as well as the knowledge of the outcome of the prediction.

Design.—Each *S* was observed under both conditions of reinforcement. The 24 *S*s were randomly assigned to two groups of equal size, the random assignment determining the first reinforcement condition under which *S* would be observed. Those *S*s observed first under low reward constituted Group L; *S*s observed first under high reward constituted Group H.

There were 100 trials in the two-choice situation in the first condition. The series of events was randomized over blocks of 20 trials with π, the probability of the more frequent event, equal to .75. The same series was used for all *S*s, with half the *S*s in each group being randomly assigned left as the position of the more frequent stimulus and the other half assigned right.

Procedure.—Each *S* was seen individually in his home for one session. In each case *S* had been told that two young women were coming to his home to "play some games" with him. The *S* and *E* sat facing one another at a small table. An assistant sat to one side of *E*.

The *E* began by explaining to *S* that he was to play a guessing game, using the materials placed before him on the table. The basic apparatus was a wooden box, 10 × 14½ in., covered at the end placed nearer *E* by a wooden platform. Attached to the platform were two metal rings, about 4½ in. apart and equidistant from the end of the box, into which were set two identical blue toy milk bottles, about 4½ in. in height.

Each *S* in Group L was then told that *E* would take away the bottles, "hide" a button in one of them, and return them to their positions on the platform, at which time *S* should "guess" where the button was, indicating his prediction by pointing. Instructions for Group H were essentially the same except for the reference to the contents of the bottles. Each *S* in this group was told that *E* would place a "prize" in one of the bottles; it was emphasized to *S* that the prize would be his to keep if he guessed its location correctly. There were two practice trials.

The *E* "hid" the button, under low reward conditions, by putting the bottles behind her back for each trial and dropping the button into one of them, according to the prearranged random series. Under high reward conditions, *E* held on her lap a box of assorted prizes hidden from *S*'s view by the lid of the box, set perpendicular to the bottom of the box and behind which the bottles could be filled. The assistant recorded the responses.

For each trial *S* indicated his prediction by pointing to one of the bottles. Regardless of the correctness of the response, *E* turned the chosen bottle upside down over the part of the box nearer *S*, where the contents, if any, fell. The bottle was then returned to its former position, and *S* was asked "Which one was it in?" to which he responded by pointing.

After the 100 trials in the initial condition, *Ss* in Group L were given the same series of 100 trials under high reward conditions. This time they were told that they were to play another game which would be the same as the first except that there would now be a prize in one of the bottles which would be theirs to keep if they guessed its location correctly. The *Ss* in Group H were given only 12 additional trials under low reward, since, as was expected, these *Ss* were usually reluctant to continue, without reward, an activity for which they had previously received prizes. These *Ss* were told that they were to play another game, the same as the first game except that now there would be a button in one of the bottles each time.

Rest periods were provided for both groups after 60 and 100 trials; *Ss* in Group L rested, in addition, after 160 trials. Additional "breaks" were provided if requested by *S* or deemed advisable by *E*. A session usually lasted from 1 to 1¼ hr.

Results and Discussion

The means of the response probabilities for all trial blocks for both groups are presented in Table 1. Asymptotic response probability, p_∞, is defined for a 100-trial series as the proportion of predictions of the more frequent event for the last 20 trials; for the 12-trial series under low reward for Group H, the proportion for the complete series is tentatively considered asymptotic response probability.

TABLE 1

Mean Response Proportions for 20-Trial Blocks

Group	Reward Condition	Trials	Trial Blocks				
			I	II	III	IV	V
L	Low	1–100	.58	.67	.59	.65	.75
	High	101–200	.72	.75	.78	.82	.85
H	High	1–100	.52	.76	.72	.78	.83
	Low	101–112	.73				

The hypothesis that preschool *Ss* observed under conditions of high reward predict the occurrence of the more frequent of two events a larger proportion of the time than do those observed under low reward conditions, thereby showing a greater tendency to maximize the probability of a correct response, was tested as follows.

First test.—Since all *Ss* were observed under both conditions of reinforcement, with counterbalancing of the order of conditions, a difference-score comparison of asymptotic response probabilities under low reward and high reward may be made. As predicted, the mean asymptotic response probability for high reward is higher than that for low reward: $\bar{p}_\infty(\mathrm{H}) = .84$ and $\bar{p}_\infty(\mathrm{L}) = .74$. A Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) shows that the difference is significant ($P < .001$).

It was recognized by the authors that a possible explanation of the higher asymptotes for those *Ss* for whom high reward was the second condition of reinforcement is that

additional learning was taking place under that condition. However, this seems not to be the case: for those *Ss* who had high reward as the first condition, $\bar{p}_\infty$ for high reward was .83; for those *Ss* who had high reward as the second condition, $\bar{p}_\infty$ for high reward was .85. A sum of ranks test (Siegel & Tukey, 1960) revealed that this difference is not significant ($P > .49$).

Second test.—Since the difference-score analysis was partially based on the 12-trial response series for the group that had low reward second, an additional comparison of asymptotes for the initial conditions of the two independent groups was made. As predicted, the mean asymptotic response probability was higher for Group H than for Group L at the end of the first 100 trials: $\bar{p}_\infty(\mathrm{H}) = .83$ and $\bar{p}_\infty(\mathrm{L}) = .75$. The differences between groups for the initial condition of reinforcement allow rejection of the null hypothesis at $.10 > P > .05$, as determined by a one-tailed sum of ranks test.[4]

Discussion

The prediction of the Siegel model is concerned specifically with asymptotic response probabilities in the two-choice situation. It is necessary, then, in order for the results of the present study to be interpreted as supporting the model, for *Ss* to have exhibited stable state behavior in the experimental sessions. As was mentioned in connection with the first test of the hypothesis, differences between p's for the last 20 trials when high reward occurred first (Group H) and the last 20 trials when high reward occurred second (Group L) are not significant. The fact that response proportions under a particular reinforcement condition for Trials 81–100 are similar to response proportions under the same reinforcement condition for Trials 181–200 is evidence in support of the use of the last 20 trials of a 100-trial series as an index of asymptotic behavior.

It may be concluded, then, that the hypothesis of the Siegel model, that asymptotic response probability is largely a function of amount of reinforcement, is supported by the results of the present study.

Stevenson and Weir (1959) have discussed the hypothesis that an increase in the incentive value of reinforcement leads to a tendency in *Ss* to seek a solution yielding consistent reinforcement. In terms of the present study, this hypothesis would lead to the prediction that an increase in amount of reinforcement leads to greater response variability, and thus a smaller proportion of predictions of the more frequent event. Since the opposite occurred, it may be concluded that the present study would not lend support to such an hypothesis.

On the basis of the apparent applicability of the Siegel model to the behavior of young children, it is suggested that an "understanding of probability" greater than that attributed to them by Piaget exists among 4-yr-olds. Piaget (1950) reports that, for the probability task he presents to young *Ss*, responses which have maximum probability of being correct seem to occur only rarely among children under 7 yr. of age. The results of the present study, however, hardly seem consistent with such findings. Although the behavior of the children under low reward conditions may be discussed without invoking the notion of a concept of probability, it would seem that the tendency to maximize probability of reinforcement under high reward conditions is, in terms of the criterion used by Piaget (1950), behavior reflecting a tendency to respond

[4] The behavior of 4 *Ss*, 2 in each experimental group, seemed to indicate an incomplete grasp of the task; asymptotes for these *Ss* were .55 or below. With these *Ss* omitted in the second test of the hypothesis, differences between groups are more significant ($P < .05$). All P values reported in the text are based on all *Ss*.

to actual "quantitative proportions" of events and hence, on the basis of this criterion, is indicative of at least a rudimentary concept of probability in the children exhibiting it. It is possible that the introduction of rewards into a probabilistic task has led to behavior which may be considered more nearly indicative of children's "understanding of probability" than performance in similar tasks in which no reinforcement is provided. Direct support of such an inference has been reported by Yost, Siegel, and Andrews (1962).

Summary

The classical two-choice uncertain outcome experimental paradigm has been adapted for use with preschool children in a test of the predictions of two models for choice behavior in the situation. Children exhibited a greater tendency to maximize the probability of a correct response as the amount of reinforcement for such a response was increased, thereby lending support to one of the alternative models under consideration. It is concluded that the behavior of children in the two-choice situation is a function of the level of amount of reinforcement for a correct response present in the situation, as well as of the event probabilities.

In addition, it is suggested that the maximizing tendency observed among very young children may be indicative of a more advanced stage in the development of a "concept of probability" in children of this age than that which may be inferred from Piaget's results.

References

Atkinson, R. C., Sommer, G. R., & Sterman, M. B. Decision making by children as a function of amount of reinforcement. *Psychol. Rep.*, 1960, **6**, 299–306.

Bijou, S. W., & Sturges, P. T. Positive reinforcers for experimental studies with children: Consumables and manipulatables. *Child Develpm.*, 1959, **30**, 151–170.

Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

Messick, S. J., & Solley, C. M. Probability learning in children: Some exploratory studies. *J. genet. Psychol.*, 1957, **90**, 23–32.

Piaget, J. Une experience sur la psychologie du hasard chez l'enfant: Le tirage au sort des couples. *Acta psychol., Amst.*, 1950, **7**, 323–336.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

Siegel, S. Decision making and learning under varying conditions of reinforcement. *Ann. N. Y. Acad. Sci.*, 1961, **89**, 766–783.

Siegel, S., & Goldstein, D. A. Decision-making behavior in a two-choice uncertain outcome situation. *J. exp. Psychol.*, 1959, **57**, 37–42.

Siegel, S., & McMichael, J. E. Individual choice and strategy behavior: A general model for repeated choices. (Res. Bull. No. 11) University Park, Pa.: Department of Psychology, Pennsylvania State University, December 1960.

Siegel, S., & Tukey, J. W. A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Statist. Ass.*, 1960, **55**, 429–445.

Stevenson, H. W., & Weir, M. W. Variables affecting children's performance in a probability learning task. *J. exp. Psychol.*, 1959, **57**, 403–412.

Yost, P. A., Siegel, A. E., & Andrews, J. M. Nonverbal probability judgments by young children. *Child Develpm.*, 1962, in press.

(Received December 16, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 342–352

STUDIES IN ASSOCIATIVE INTERFERENCE[1]

NORMA F. BESCH, VENAN E. THOMPSON, AND ALLAN B. WETZEL

Columbus Psychiatric Institute and Hospital, Ohio State University

Twedt and Underwood (1959) have made a distinction between the experimental designs used in the study of verbal proactive transfer in terms of the types of lists used. In the design using "mixed lists," different relationships hold between subsets of items in Lists 1 and 2. Thus, with one group of *S*s learning only two lists, two or more transfer paradigms are being investigated simultaneously. In "unmixed lists," the same relationship holds between all items in Lists 1 and 2 or one paradigm is represented in the two lists. Thus, in a design using unmixed lists, in which two or more proactive paradigms are being compared, each group learns a different set of two lists.

Using mixed lists, Besch and Reynolds (1958) found negative transfer in two paradigms when they were compared to an A-B, D-C control paradigm (the stimuli and responses are both different). The negative transfer paradigms were: (*a*) A-B, A-C (the stimuli are the same but the responses are different), and (*b*) A-B, A-Br (the stimuli and responses are the same but are paired differently). It was suggested that the increase in the negative transfer found for the A-B, A-C paradigm over that usually found with unmixed lists was not due to the greater precision gained with the use of mixed lists, but to the mixing itself. Generalization of habit strength was the means by which the increase was to have occurred. In brief, by the occurrence of stimulus generalization within the first list, associative tendencies were established between the stimuli of the A-B, A-C paradigm and the responses of the A-B, A-Br paradigm. These tendencies, which were further strengthened by generalization during transfer task learning, competed effectively with the correct tendencies, the A-C pairings, on the transfer task.

For the A-B, A-Br paradigm, it was assumed that the associative tendencies established in List 1 were being further strengthened by generalization during the transfer task (List 2). These associative tendencies competed effectively during transfer task learning with the correct tendencies, the A-Br pairings. However, the A-Br pairings, unlike the A-C pairings, were of some strength at the beginning of List-2 learning due to the effects of generalization during List-1 learning. By this interpretation, negative transfer for the A-B, A-Br paradigm would not be a function of list type, since the relationships believed to be critical are the same in both mixed and unmixed lists.

The purpose in the present experiment was to re-examine the generalization interpretation of associative interference by studying pair types intermediate to the A-Br and A-C transfer pair types. Two features distinguish the A-Br from the A-C pair: (*a*) the stimulus of an A-Br pair is paired with a List-1 response which was incorrect for that stimulus on List 1, and (*b*) the response, which was paired with the stimulus of the A-Br pair in List 1, reappears in the transfer task and is paired with another stimulus from List 1. Thus, one intermediate pair type will be formed by pairing a List-1 stimulus with the response term of another List-1 pair. The other type

[1] This research was supported in part by Grant G-9061 from the National Science Foundation.

will be formed by pairing a new response with a stimulus from a List-1 pair; the response term of that List-1 pair will be paired with a stimulus from a different List-1 pair. The latter paradigm will be represented as A_1-B_1, A_1-C and the former as A_2-B_2, A_2-B_1.

Using mixed lists, the A_1-C, A_2-B_1, A-Br, and D-C transfer pair types will be compared in Exp. 1, and the A_1-C, A_2-B_1, A-C, and D-C transfer pair types will be compared in Exp. 2 and 3.

Experiment 1

Method

Subjects.—The *Ss* were 40 female volunteers from the introductory psychology course at Ohio State University. They were assigned in equal numbers of one of four counterbalancing groups. An attempt was made in the assignment of *Ss* to obtain approximately equal mean scores for the different groups on a practice list.

Design.—The *Ss* in each group learned in order a practice list, a pretraining list, and two experimental lists, Lists A and B. Lists A and B were the lists representing the proactive paradigm. The pretraining list was used in an attempt to minimize the effects of word familiarity on the learning of List B, the transfer task. The pretrained words were those words which were present in List B and not present in List A. Since there were four forms of List A, there were also four forms of the pretraining list.

List B, the transfer task, which was identical for all groups, is best considered as consisting of four sets of three word pairs each. The four forms of List A, each learned by a different group of *Ss*, were constructed to allow each of these four sets of three pairs to serve in turn under each of the transfer task conditions. The relationships between the word pairs of the four forms of List A and those of List B are presented in Table 1. The relationship between any pair set within Form A and the corresponding transfer pair set, which appears in the last column of the same row, is represented in each case by the paradigm name. The four paradigms are: (*a*) A-B, D-C (three of the pairs of List A were replaced by new S and R terms), (*b*) A-B, A-Br (the S and R terms from three of the List-A pairs were repaired), (*c*) A_2-B_2, A_2-B_1 (three of the stimuli from List A were paired with the R terms from three other List-A pairs), and (*d*) A_1-B_1, A_1-C (new R words were paired in the transfer task with stimuli from three other List-A pairs, the A_2-B_2 pairs). To see clearly the relationship between the A_1-B_1, A_1-C, and the A_2-B_2, A_2-B_1 paradigms, it is necessary to examine in Table 1 more than the sets within the same row. For any A_2-B_2 set in List A, the response words of the corresponding transfer pair set are the response words of the A_1-B_1 set in List A.

Lists and apparatus.—The practice list contained 15 paired one-syllable nouns of low associative strength. List B and the four forms of List A each consisted of 12 pairs of two-syllable adjectives from the Haagen (1949) list. The word pairs for the transfer task were chosen in a way which, in the writer's opinion, minimized meaningful associations, synonymities, and formal similarities both within and between pairs. Thus, List A pairs having words in common with the transfer task met these same conditions. As shown in Table 1, the List A word pairs serving under the A-B, D-C condition and the List A response words serving under the A_2-B_2, A_2-B_1 condition were identical in the four forms of List A. These words were again chosen to satisfy the requirements for the transfer task.

Each counterbalancing group, which learned a different form of List A, also learned a different form of the pretraining list. The response words of each pretraining list consisted of all words which were present in List B and not present in the appropriate form of List A. The stimulus terms of each pretraining list were the first two letters of their paired response words.

A Hunter Model 340 Card Master was used to present all tasks. The 6-sec. interstimulus interval included a 2-sec. stimulus exposure, a 2-sec. stimulus and response exposure, and a 2-sec. blank. A 6-sec. rest interval was given between successive presentations of a list. Three presentation orders were used for each list.

Procedure.—Each *S* learned by the anticipation method four lists in the following order: (*a*) a practice list, (*b*) a pretraining list, (*c*) List A, and (*d*) List B. Six trials each were given on both the practice list and the pretraining list. All *Ss* received 12 trials on List A. A criterion of 2 consecutive errorless trials was set for the transfer task; however, if *S* had not reached criterion by Trial 15, practice was discontinued. In the analyses of the data, the assumption was made that if *S*,

TABLE 1

DESIGN FOR EXPERIMENT 1

List A				List B
Form 1 (Group 1)	Form 2 (Group 2)	Form 3 (Group 3)	Form 4 (Group 4)	(All Groups)
A-B, D-C	A-B, A-Br	A_2-B_2, A_2-B_1	A_1-B_1, A_1-C	
S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_1—R_2 S_2—R_3 S_3—R_1	S_1—R_{16} S_2—R_{17} S_3—R_{18}	S_1—R_4 S_2—R_5 S_3—R_6	S_1—R_1 S_2—R_2 S_3—R_3
A-B, A-Br	A-B, D-C	A_1-B_1, A_1-C	A_2-B_2, A_2-B_1	
S_4—R_5 S_5—R_6 S_6—R_4	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_4—R_1 S_5—R_2 S_6—R_3	S_4—R_{16} S_5—R_{17} S_6—R_{18}	S_4—R_4 S_5—R_5 S_6—R_6
A_2-B_2, A_2-B_1	A_1-B_1, A_1-C	A-B, D-C	A-B, A-Br	
S_7—R_{16} S_8—R_{17} S_9—R_{18}	S_7—R_{10} S_8—R_{11} S_9—R_{12}	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_7—R_8 S_8—R_9 S_9—R_7	S_7—R_7 S_8—R_8 S_9—R_9
A_1-B_1, A_1-C	A_2-B_2, A_2-B_1	A-B, A-Br	A-B, D-C	
S_{10}—R_7 S_{11}—R_8 S_{12}—R_9	S_{10}—R_{16} S_{11}—R_{17} S_{12}—R_{18}	S_{10}—R_{11} S_{11}—R_{12} S_{12}—R_{10}	S_{13}—R_{13} S_{14}—R_{14} S_{15}—R_{15}	S_{10}—R_{10} S_{11}—R_{11} S_{12}—R_{12}

who had reached criterion, had continued practice, his performance would have been errorless for the remainder of the 15 trials. Ten of the 40 *Ss* reached criterion before Trial 15; all of these *Ss* reached the criterion between Trials 10 and 15.

Standard instructions for paired-associate learning were given before the presentation of the practice list. The *Ss* were informed before each succeeding task that the same procedure was to be followed in learning the new list. The *Ss* were not informed concerning the relationship of the pretraining list to the other lists. The time interval between tasks was approximately 3 min.

Results

A simple analysis of variance of the number of correct responses on List A yielded an F of less than 1.00 for the main effect of lists. In addition, an analysis of variance of correct responses for the List-A pairs which were identical in the four forms, that is, the pairs serving under the A-B, D-C condition, gave the same result.

Figure 1 shows the percentages of correct responses on each transfer trial for each pair type. The analysis of variance (Lindquist, 1953, Type VI) of the number of correct responses for all groups over the 14 anticipation trials is summarized in Table 2. Due to the significant interactions between

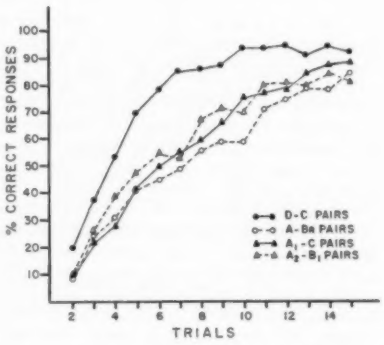

FIG. 1. Percentage of correct responses in Exp. 1 for pair conditions in Trials 2–15 of transfer task.

TABLE 2

ANALYSIS OF VARIANCE OF CORRECT RESPONSES FOR PAIR CONDITIONS ON TRIALS 2–15 OF THE TRANSFER TASK: EXP. 1

Source	*df*	*MS*	*F*
Total between *Ss*	39	11.68	
Groups (G)	3	2.60	.21
Error (between)	36	12.44	
Total within *Ss*	2200	1.04	
Pair conditions (P)	3	51.14	23.57***
Trials (T)	13	81.10	193.10***
P × T	39	.84	2.33***
P × G	9	5.91	2.72**
T × G	39	.44	1.05
P × T × G	117	.45	1.25
Residual	1980	.47	
$Error_1$	108	2.17	
$Error_2$	468	.42	
$Error_4$	1404	.36	

** $P < .01$.
*** $P < .001$.

Pairs and Trials and between Pairs and Groups, *t* tests of the pair differences were performed for Groups and Trials separately. Performance under Cond. D-C was significantly superior ($P < .05$) to Cond. A-Br on Trials 4–11, to Cond. A_1-C on Trials 4–9, and to Cond. A_2-B_1 on Trials 5, 6, 7, and 10; no other differences were significant. The means and *SD*s of the performance for each pair condition for the four counterbalancing groups are shown in Table 3. The means for Cond. D-C were significantly larger ($P < .05$) than those for all other pairs in all but three comparisons, which were D-C vs. A_1-C and D-C vs. A_2-B_1 for Group 3 and D-C vs. A-Br for Group 2. In addition, for Group 2 the A_2-B_1 mean was significantly smaller than the A-Br mean, and for Group 3 the A_2-B_1 mean was significantly larger than the A-Br mean. Thus, it may be concluded tentatively that in comparison to the D-C pairs, all three of the other pair conditions show negative transfer, at least for the middle portion of learning.

Finally, the intruding responses on the transfer task for each pair condition are classified by source in Table 4. Only a few stimulus words intruded; these were placed in the "Other" category. The number of intruding responses for the pair conditions appears to follow inversely the order of correct responses on the pairs. However, it should be noted that while the difference between the number of intrusions for the A_2-B_1 and A_1-C transfer pairs was large, the difference for correct responses was small.

EXPERIMENT 2

Method

Subjects and design.—Forty female *Ss* were selected, assigned to groups, and given practice on four lists in the same manner as in Exp. 1. In Exp. 2, the A-B, A-C paradigm replaced the A-B, A-Br paradigm of Exp. 1. This was accomplished by selecting three new response words to replace the response words which were used in Exp. 1 in the pair sets of the four forms of List A which served under Cond. A-B, A-Br.

Lists, apparatus, and procedure.—The practice list and the transfer list of Exp. 2

TABLE 3

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES ON TRANSFER TASK IN EXP. 1

Groups	Pair Condition							
	D-C		A-Br		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	35.2	4.4	23.7	8.8	26.5	7.4	25.5	8.1
2	32.3	6.7	27.5	7.0	24.5	8.6	22.6	9.3
3	28.9	5.6	19.6	8.0	24.4	5.0	29.2	10.5
4	34.1	6.0	20.9	10.2	24.4	7.6	24.7	7.2

TABLE 4

SOURCE AND NUMBER OF INTRUDING RESPONSES ON THE TRANSFER TASK OF EXP. 1 FOR EACH PAIR CONDITION

Source	Transfer Pair Condition			
	D-C	A-Br	A_1-C	A_2-B_1
A-B, D-C				
List B	26	13	12	13
A-B, A-Br				
Correct (List A)	—	109	—	—
Incorrect (List A)	10	8	13	16
A_1-B_1, A_1-C				
Correct (List A)	—	—	90	—
Incorrect (List A)	—	—	22	—
List B	15	9	9	23
A_2-B_2, A_2-B_1				
Correct (List A)	—	—	—	7
Incorrect (List A)	1	1	4	0
List B	10	15	—	28
Other	12	13	9	9
Unidentified	1	8	10	10
Total	75	176	169	106
Mean	1.9	4.4	4.2	2.6
SD	2.2	3.2	2.4	2.4

Note.—$N = 40$.

were identical to those of Exp. 1. With the exception of one pair set in each form, the four forms of List A were the same in both experiments. In Exp. 2, the same three response words were used in all forms of List A to replace the response words of the pair sets of Cond. A-B, A-Br in Exp. 1. Again, these response words were chosen to minimize, in the four forms of List A, meaningful associations, synonymities, and formal similarities both within and between pairs.

Since all words not present in List A and present in List B were used to construct the pretraining lists, three new pairs were added to the pretraining list of Exp. 1 to construct the lists for Exp. 2.

The apparatus and the procedures used in Exp. 1 were unchanged in Exp. 2.

Twenty-one of the 40 *S*s reached criterion on the transfer task before Trial 15; of these *S*s, 6 reached criterion before Trial 10 and the remainder after Trial 10.

Results

Simple analyses of variance of List-A correct responses and of correct responses on the List-A pairs of the A-B, D-C pair sequence yielded *F*s of 1.52 ($P > .20$) and 1.18 ($P > .20$), respectively, for the main effect of Lists.

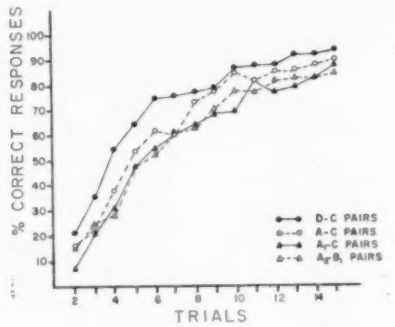

FIG. 2. Percentage of correct responses in Exp. 2 for pair conditions in Trials 2–15 of transfer task.

Figure 2 shows the percentages of correct responses on each transfer trial for each pair type. The analysis of variance of the number of correct responses for all groups over the 14 anticipation trials is summarized in Table 5. Due to the significant interaction between Pairs and Groups, *t* tests were performed for the pair differences within each group. The means and *SD*s of the correct responses for the transfer pair conditions for each of the four groups are pre-

TABLE 5

ANALYSIS OF VARIANCE OF CORRECT RESPONSES FOR PAIR CONDITIONS ON TRIALS 2–15 OF THE TRANSFER TASK: EXP. 2

Source	*df*	*MS*	*F*
Total between *S*s	39	18.25	
Groups (G)	3	26.95	1.54
Error (between)	36	17.52	
Total within *S*s	2200	.91	
Pair conditions (P)	3	18.20	18.76***
Trials (T)	13	79.54	162.33***
P × T	39	.39	1.18
P × G	9	3.54	3.65***
T × G	39	.80	1.63*
P × T × G	117	.37	1.12
Residual	1980	.40	
$Error_1$	108	.97	
$Error_2$	468	.49	
$Error_3$	1404	.33	

* $P < .05$.
*** $P < .001$.

TABLE 6

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES ON TRANSFER TASK IN EXP. 2

Groups	Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	33.3	7.8	25.2	8.0	22.2	10.3	22.5	4.6
2	32.2	7.9	33.9	4.2	28.5	10.2	28.8	5.5
3	26.5	7.6	23.4	6.7	21.8	9.6	24.2	6.0
4	31.1	12.1	28.6	7.9	28.6	9.1	27.1	6.2

sented in Table 6. In comparison to the D-C pairs, the A_1-C and the A_2-B_1 pairs show negative transfer in three out of four of the comparisons. The mean differences were not significant for D-C vs. A_2-B_1 and D-C vs. A_1-C for Groups 3 and 4, respectively. Only in Group 1 was a significant *t* obtained for A-C vs. D-C. Only one A-C vs. A_1-C and one A-C vs. A_2-B_1 comparison were significant; both in Group 2. The significant Trials × Groups interaction indicates that the groups approach their asymptotes at different rates. Since there is little interest in this finding, *t* tests of the group differences over trials were not performed.

The intruding responses on the transfer task for each pair condition are classified by source in Table 7. Again, intruding stimulus words are placed in the "Other" category. As in Exp. 1, the number of intruding responses is inversely related to the number of correct responses for each pair condition. Again the differences were not of comparable magnitudes for intrusions and correct responses.

TABLE 7

SOURCE AND NUMBER OF INTRUDING RESPONSES ON THE TRANSFER TASK OF EXP. 2 FOR EACH PAIR CONDITION

Source	Transfer Pair Condition			
	D-C	A-C	A_1-C	A_2-B_1
A-B, D-C				
List B	15	21	7	12
A-B, A-C				
Correct (List A)	—	5	—	—
Incorrect (List A)	0	1	0	1
List B	33	4	8	23
A_1-B_1, A_1-C				
Correct (List A)	—	—	91	—
Incorrect (List A)	—	—	27	—
List B	19	23	11	36
A_2-B_2, A_2-B_1				
Correct (List A)	—	—	—	4
Incorrect (List A)	3	2	1	1
List B	1	22	—	13
Other	17	9	21	5
Unidentified	0	8	6	5
Total	88	95	172	100
Mean	2.2	2.4	4.3	2.5
SD	2.5	2.4	3.9	2.6

EXPERIMENT 3

The pretraining procedures used in Exp. 1 and 2 may not have affected the different pair types in the same way. Therefore, in Exp. 3, Exp. 2 was replicated over two conditions of pretraining, one relevant and the other irrelevant.

Method

Subjects.—The *S*s were 48 women volunteers from the introductory psychology course at Ohio State University. They were assigned in equal numbers to one of eight groups. Again, the assignment of *S*s to groups was made on the basis of practice list scores.

Design and lists.—The four groups, which served in the relevant pretraining condition, were trained in the same way and learned the same lists as *S*s in Exp. 2. The four

irrelevant pretraining groups differed in one aspect from the other groups; the words of the pretraining lists did not appear in any of the subsequent experimental lists. For the four forms of the irrelevant pretraining lists, words were used with the same beginning letters as the words used for the comparable relevant pretraining lists. Where there were repetitions in the words used in the different relevant pretraining lists, repetitions were likewise present across the lists for the irrelevant pretraining groups. Again, the words were selected from the Haagen (1949) list to minimize meaningful associations and synonymities both within the pretraining lists and between it and the other lists.

Apparatus and procedure.—The Lafayette memory drum, Model 303A, which was used to present the practice and transfer lists, operated by exposing with each turn a stimulus word or a stimulus and response word together. The Lafayette Model 303B, which was used to present the remainder of the lists, had an automatic shutter which alternately covered and uncovered the response word of a pair. In the use of both drums, the interstimulus interval was 4 sec. as opposed to the 6-sec. interval of Exp. 1 and 2. The Lafayette drums do not include a blank between pairs as does the Hunter Card Master.

All other procedures used in Exp. 2 were unchanged in Exp. 3.

Of the 11 *Ss* in the relevant pretraining groups who reached criterion earlier than Trial 15, 4 reached it before and 7 after Trial 10. For the 7 *Ss* of the irrelevant pretraining groups reaching an early criterion, 2 reached it before and 5 after Trial 10.

Results

A factorial analysis of variance of the number of correct responses on List A yielded nonsignificant *F*s for the main effect of Lists, the main effect of Pretraining Groups, and the Groups × Lists interaction. The factorial analysis of correct responses on the List-A pairs serving under Cond. A-B, D-C also yielded nonsignificant *F*s.

Figures 3 and 4 present, respectively, for the relevant and the irrelevant pretraining groups the percentages of correct responses on each transfer trial for each pair type.

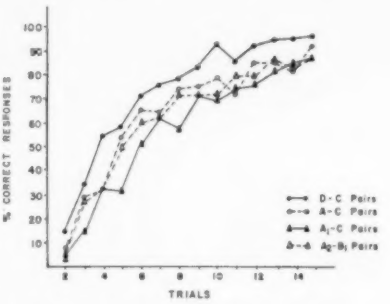

FIG. 3. Percentage of correct responses for pair conditions in Trials 2–15 of transfer task for relevant pretraining groups of Exp. 3.

Table 8 presents the means and *SD*s of the correct responses for the transfer pair conditions for each of the counterbalancing groups of both the irrelevant and the relevant pretraining conditions.

For the relevant pretraining groups, analysis of variance (Lindquist, 1953, Type VI) yielded results similar to those found in Exp. 2; the main effects of Pairs and Trials were significant ($P < .001$), but only the Pairs × Groups interaction was significant ($P < .05$). The t tests, warranted by the significant interaction, gave results different from those of Exp. 2 in only two comparisons;

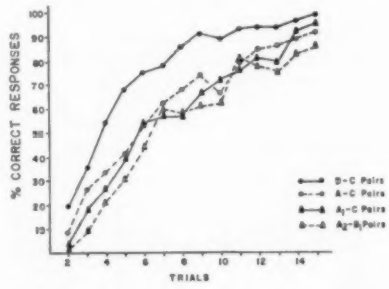

FIG. 4. Percentage of correct responses for pair conditions in Trials 2–15 of transfer task for irrelevant pretraining groups of Exp. 3.

four rather than three of the A_1-C vs. D-C comparisons were significant, and one rather than none of the A_1-C vs. A_2-B_1 comparisons, that for Group 3, was significant. However, the single significant t's for the D-C vs. A-C, A-C vs. A_1-C, and A-C vs. A_2-B_1 comparisons were obtained for Groups 4, 3, and 1, respectively.

For the irrelevant pretraining groups, the same type of analysis of variance yielded significant main effects only, that for Pairs ($P < .001$) and Trials ($P < .001$). The t tests for pairs, ignoring counterbalancing groups, indicated that all three experimental pairs differed significantly from the control (D-C) pairs.

From an inspection of Fig. 3 and 4, it appears that the type of pretraining does affect pair performance differently on the transfer task. Thus, the two pretraining conditions were compared with each other in an analysis of variance which ignored the trials factor. An analysis of variance (Lindquist, 1953, Type III) of the total number of correct responses for each pair type for each group was performed. Even though the relevant pretraining groups showed better performance on the experimental pairs and poorer performance on the D-C control pairs in comparison to the irrelevant pretraining groups, the Pairs × Pretraining Groups interaction was not significant ($F = 2.08$, $P < .20$). Only the main effect of Pairs and the Pairs × Counterbalancing Groups interactions were significant ($P = .001$ and .05, respectively).

The intruding responses by pair source for the relevant and irrelevant pretraining groups are presented in Table 9. The intruding responses did not relate inversely to the correct responses as in Exp. 2. The same order for correct responses held for the relevant pretraining groups, but the order of intrusions was changed by the A_2-B_1 pairs which had very few intrusions, in fact, fewer than the D-C pairs. Again, performance was poorest for the A_1-C pairs, along with the largest number of intrusions. In the irrelevant pretraining conditions,

TABLE 8

MEANS AND *SD*s OF TOTAL CORRECT RESPONSES ON TRANSFER TASK IN EXP. 3

Groups	Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Relevant Pretraining								
1	31.8	3.7	29.5	3.6	25.5	7.0	24.2	7.1
2	27.7	12.7	22.8	14.4	22.0	10.7	22.2	10.8
3	31.3	5.6	32.0	5.3	23.2	9.3	32.2	5.7
4	32.5	6.5	23.7	8.3	25.3	9.5	25.8	7.4
Irrelevant Pretraining								
1	32.7	7.2	24.8	5.9	23.0	6.1	17.8	7.8
2	34.2	5.3	26.0	9.5	23.3	7.6	22.3	7.5
3	33.2	5.8	27.0	8.5	22.0	9.3	28.7	6.9
4	28.8	4.2	25.5	9.5	29.5	4.7	21.7	7.6

TABLE 9

SOURCE AND NUMBER OF INTRUDING RESPONSES ON THE TRANSFER TASK FOR EACH PAIR CONDITION FOR THE RELEVANT AND IRRELEVANT PRETRAINING GROUPS

Source	Transfer Pair Condition							
	D-C		A-C		A_1-C		A_2-B_1	
	R[a]	I[b]	R[a]	I[b]	R[a]	I[b]	R[a]	I[b]
A-B, D-C								
List B	7	15	11	2	5	0	5	3
A-B, A-C								
Correct (List A)	—	—	1	1	—	—	—	—
Incorrect (List A)	0	0	0	0	1	0	0	0
List B	16	6	9	9	5	10	8	0
A_1-B_1, A_1-C								
Correct (List A)	—	—	—	—	55	59	—	—
Incorrect (List A)	—	—	—	—	12	6	—	—
List B	5	3	14	12	10	5	8	13
A_2-B_2, A_2-B_1								
Correct (List A)	—	—	—	—	—	—	3	6
Incorrect (List A)	0	0	0	1	0	0	1	0
List B	7	2	8	8	—	—	6	12
Other	6	1	5	2	4	1	2	1
Unidentified	3	1	4	2	6	4	4	3
Total	44	28	52	37	98	85	37	38
Mean	1.8	1.2	2.2	1.5	4.1	3.5	1.5	1.6
SD	2.9	2.0	3.1	2.1	4.3	3.3	2.2	1.6

Note.—$N = 24$.
[a] R refers to relevant pretraining groups.
[b] I refers to irrelevant pretraining groups.

for the first time performance was poorer for the A_2-B_1 pairs in comparison to the A_1-C pairs, but they did not surpass the A_1-C pairs in the number of intruding responses. The number of intrusions was lower for each pair type, with the exception of the A_2-B_1 pairs, for the irrelevant in comparison to the relevant pretraining condition. One would expect the A_2-B_1 pairs to be unaffected since they were not involved in the pretraining.

DISCUSSION

In agreement with previous findings (Besch & Reynolds, 1958; Porter & Duncan, 1953; Twedt & Underwood, 1959), negative transfer was shown for A-Br pairs. For the irrelevant pretraining groups in Exp. 3, the A-C pairs were significantly different from the D-C pairs, while in Exp. 2 and for the relevant pretraining groups in Exp 3, the A-C pairs did not differ significantly from the D-C pairs. In all of the comparisons between the A-C and the D-C pairs, the differences were in the direction shown in comparable experiments (Besch & Reynolds, 1958; Spiker & Holton, 1958; Twedt & Underwood, 1959). Since Underwood, Runquist, and Schulz (1959) showed that pretraining of the response words of a paired-associate list facilitated subsequent performance on that list, it seems probable that the A-C pairs in the present experiments were facilitated by relevant pretraining. While the difference obtained in Exp. 3 was in the appropriate direction, it was not statistically significant. Therefore, one can conclude only tentatively that the pretraining of both stimuli and responses, as in the D-C pairs, either depresses performance on the subsequent pairings, or facilitates performance to a minor

degree in comparison to the facilitation obtained with response pretraining. Finally, both the A_1-C and A_2-B_1 pairs showed negative transfer in all experiments, with the performance levels for these pairs above the level for the A-Br pairs and below that for the A-C pairs, but not significantly so in either case.

The primary purpose underlying the present experiments was to determine whether generalization is a sufficient principle in accounting for negative transfer effects. In an earlier paper (Besch & Reynolds, 1958), in which a "response competition" interpretation was offered, it was implicitly assumed that as long as the response items from the first list were not present in the transfer task (A-B, A-C paradigm), their effects were negligible and could be ignored. However, Spiker and Holton (1958) demonstrated that such is not the case; in the A-B, A-C paradigm, negative transfer, or at least a large portion of it, is due to competition from first task responses. In addition, it also seems clear from the results of Barnes and Underwood (1959) that the A-B stimulus and response associations are gradually extinguished during the course of A-C learning. As Postman (1959) suggests, any account of interference in verbal learning must encompass "unlearning of first task responses" as well as "response competition." Earlier, Melton and Irwin (1940) made this same suggestion. With these considerations in mind, an attempt will be made to evaluate the generalization interpretation of interference by comparing the A_1-C and the A_2-B_1 pairs with the A-C and the A-Br pairs.

According to a generalization interpretation, what would be expected for an A_1-C pair in comparison to an A-Br pair? The previously correct responses on List A for both pair types are present in the transfer task. By the process of generalization of habit strength, these tendencies would be gradually strengthened throughout transfer task learning and thus be maintained at a sufficient strength to compete successfully with the correct tendencies. These interfering tendencies would not be competing with correct tendencies of equal strength, however. The correct tendency for the A-Br pair would be of some strength at the beginning of transfer task learning due to the generalization of associative strength on the first task (List A), while the correct tendency for the A_1-C pair would be of zero or negligible associative strength. The A_1-C pairs, it will be recalled, were easier rather than more difficult than the A-Br pairs.

For the A_2-B_1 pair as for the A-Br pair, the correct tendency on the transfer task has some strength gained from generalization during first-task learning. The previously correct tendency is being strengthened for the A-Br pair, while it is being extinguished for the A_2-B_1 pair. The A_2-B_1 pairs, as would be predicted, were easier than the A-Br pairs.

The A_1-C and A-C pairs would not differ from each other in the strength of the correct tendencies on the transfer task. The A_1-C pairs were more difficult than the A-C pairs, as would be predicted, since the previously correct tendency is being extinguished for the A-C pair, and being strengthened for the A_1-C pair during transfer task learning.

The A_2-B_1 pair differs in one respect from the A-C pair; its correct tendency is of above zero strength due to generalization during first task learning, while that for the A-C pair is of zero or negligible strength. However, the A_2-B_1 pairs were more difficult, rather than easier, than the A-C pairs.

In summary, the direction of the differences shown in half of the comparisons was opposite to that predicted. While these differences were not statistically significant, it does seem reasonable to place confidence in them, particularly since the findings of Exp. 2 were replicated in Exp. 3. It may be concluded that the generalization interpretation, offered in the present paper, by itself does not adequately explain the findings.

It appears necessary to assume that at some point during the learning of any list, associations between nonpaired stimulus and response terms undergo extinc-

tion or inhibition. It may be recalled that Gibson (1940) found a similar assumption useful. She hypothesized that, in addition to generalization, a simultaneous process of differentiation is occurring which decreases generalization. Although Gibson did not state how generalization was decreased, she did state that differentiation in verbal learning is analogous to differential inhibition in classical conditioning. An extinction hypothesis, used alone or as used by Gibson, is consistent with the results found in the present experiments. For example, in the A_2-B_1 pair, the correct associations on Task B have been subject to extinction effects on Task A, and thus the probability is decreased that a stimulus of an A_2-B_1 pair will elicit its paired response. This is also true of the A-Br pair. For the A-Br pair, as well as the A_1-C pair, its previously correct response is present in the list; the tendency has some strength and therefore needs to be extinguished during transfer task learning. The situation in the transfer list for the A-Br pairs and the A_1-C pairs would thus be analogous to the existence of preexperimental associations in the Task A. In conclusion, it is felt that an extinction interpretation of interference has some merit, since it would predict the A-C pairs to show the least interference, the A-Br pairs to show the most interference, and the A_1-C and A_2-B_1 pairs to be intermediate to the A-C and A-Br pairs.

Summary

Three experiments were performed for the purpose of evaluating a generalization interpretation of interference. In Exp. 1, four transfer paradigms were compared: (*a*) A-B, D-C, (*b*) A-B, A-Br, (*c*) A_1-B_1, A_1-C, and (*d*) A_2-B_2, A_2-B_1. In comparison to D-C transfer pair learning, all other pairs showed negative transfer. No other differences were significant.

In Exp. 2, four transfer paradigms were compared: (*a*) A-B, D-C, (*b*) A-B, A-C, (*c*) A_1-B_1, A_1-C, and (*d*) A_2-B_2, A_2-B_1. Only the A_1-C and A_2-B_1 pairs showed negative transfer.

In Exp. 3, Exp. 2 was replicated over two conditions of pretraining, one relevant and the other irrelevant. Only the A_1-C and A_2-B_1 pairs showed negative transfer for the relevant pretraining conditions. All three experimental pair types showed negative transfer for the irrelevant pretraining condition.

The generalization interpretation which was presented could not account for the findings. An alternative explanation in terms of extinction or unlearning was suggested.

References

Barnes, J. M., & Underwood, B. J. Fate of first-list associations in transfer theory. *J. exp. Psychol.*, 1959, **58,** 97–105.

Besch, N. F., & Reynolds, W. F. Associative interference in verbal paired-associate learning. *J. exp. Psychol.*, 1958, **55,** 554–558.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47,** 196–229.

Haagen, C. H. Synonymity, vividness, familiarity, and association value ratings of 400 pairs of common adjectives. *J. Psychol.*, 1949, **27,** 453–463.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

Melton, A. W., & Irwin, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.

Porter, L. W., & Duncan, C. P. Negative transfer in verbal learning. *J. exp. Psychol.*, 1953, **46,** 61–64.

Postman, L. The present status of interference theory. Paper read at Conference on Verbal Learning, Ardsley-On-Hudson, Nov. 1959.

Spiker, C. C., & Holton, R. B. Associative transfer in motor paired-associate learning as a function of amount of first-task practice. *J. exp. Psychol.*, 1958, **56,** 123–132.

Twedt, H. M., & Underwood, B. J. Mixed versus unmixed lists in transfer studies. *J. exp. Psychol.*, 1959, **58,** 111–116.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. *J. exp. Psychol.*, 1959, **58,** 70–78.

(Received December 27, 1960)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 353–360

BRIGHTNESS DISCRIMINATIONS WITH CONSTANT DURATION INTERMITTENT FLASHES[1]

ROBERT L. ERDMANN

IBM Corporation

Under certain conditions intermittent stimulation is more effective than constant stimulation for producing brightness effects. This enhancement is referred to as the Brücke effect and is obtained with low frequency flashing lights. Bartley (1941) has obtained a maximum enhancement with intermittent stimulation rates of about 9 cycles/sec. He has related these brightness enhancement curves to the alpha rhythm of the brain and has postulated that a central neural mechanism is responsible for enhancement.

Using constant duration flashes, Schuil (1940) found a similar brightness effect for frequencies between 5–13 flashes/sec. He found that flashes at these frequencies have an apparent intensity greater than either Blondel-Rey's correction or Talbot's law would indicate. Also using conconstant duration flashes, Bartlett[2] obtained probability of brightness detection data as a function of flash frequency for rates from 1 to 20 flashes/sec. Although these data are unpublished, the curves showed a maximum detection probability at rates varying from 4 to 12 flashes/sec.

The above studies have all used low frequency intermittent stimulation to obtain a brightness effect. However, they have not been directly concerned with brightness discrimination, or they have confounded flash duration and flash frequency, or there have been special difficulties in specifying the stimulus conditions. At present, systematic data are lacking in which low frequency constant duration flashes have been used as a stimulus for studying brightness discrimination. Therefore, the present study, under conditions which permit reliable specification of the stimulus, was performed to determine the probability of detection of constant duration intermittent flashes as a function of flash frequency with flash and background luminance as parameters.

METHOD

Apparatus.—A diagram of the optical system is shown in Fig. 1. Light for both the adaptation field and the test field was provided by Sylvania R1131C glow tubes (A, A_1). For the adaptation system, light was collimated by an achromatic lens (L_1), passed through a filter holder (F_1), and area limiting field stop (FS_1), a condensing lens (L_2) and formed a Maxwellian view for *S* through a cube beam splitter (BS_2) and 2-mm. artificial pupil. The test field had a similar optical arrangement with the addition of a balanced wedge (BW) to provide a rapid adjustment of luminance.

Light for the fixation pattern was collimated (L_3), transmitted through a red filter (RF), a glass slide (GS), and added to the adaptation system by a beam splitter (BS_1). The fixation pattern consisted of four dots which, edge to edge, subtended a visual angle of 2° in the center of the adaptation field. Fixation luminance was controlled by *S* with a potentiometer and adjusted to a just visible level.

[1] Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Pure Science, Columbia University. The author wishes to express his appreciation to Conrad G. Mueller for his many helpful suggestions.

[2] Personal communication from Professor N. R. Bartlett on brightness discrimination obtained with a phosphor screen. Unexpected drops in his detection probabilities actually led to the present experiment.

Luminance was calibrated by removing the condensing lens (L_2) from the adaptation field and inserting a diffusing screen and a field stop matched in area (14°30′) and aligned with light from the test field system. Opposite halves of the two fields were blocked and a brightness match made. After turning off the test field and removing the artificial pupil, a Macbeth illuminometer was used to measure luminance. This reading provided a reference luminance in the test field, which, by use of a split field technique, was used to match the luminance of adaptation field when a Maxwellian view was provided.

Once the luminance was calibrated, a constant monitor was available with the aid of a photomultiplier tube (P) built into the apparatus. The photomultiplier circuit, basically that of Hamburger and King (1948), used a 931A photomultiplier tube corrected to approximate the response of the photopic eye by means of a Wratten 15 filter. Thus, any change in the spectral output of the glow tube affected the phototube monitor in the same manner as it would the photic response of the eye. A portion of the light from the cube beam splitter was picked up by the photomultiplier. After amplification, this appeared as a voltage deflection on a Tektronix dual channel scope (Model 545). Matching the voltage deflections with the Macbeth reading provided a monitor of the luminance throughout the experiment.

Figure 2 is a block diagram of the apparatus with associated monitoring equipment. A voltage drop across the glow modulator tube was fed into a Berkeley universal timer and counter (Model 5500) and into one channel of a dual channel Tektronix scope. The Berkeley timer could time the flash duration to the nearest 0.00001 sec. and the interval between flashes with the same accuracy. The timer, along with a calibrated sweep speed on the Tektronix, enabled accurate determination of flash frequency and provided a cross check on the calibration of the two instruments. After the frequency calibrations, the Berkely could be switched over to a counter operation for monitoring the number of flashes in each train of pulses.

To calibrate the wave form of the stimulus, the output of the photomultiplier tube was

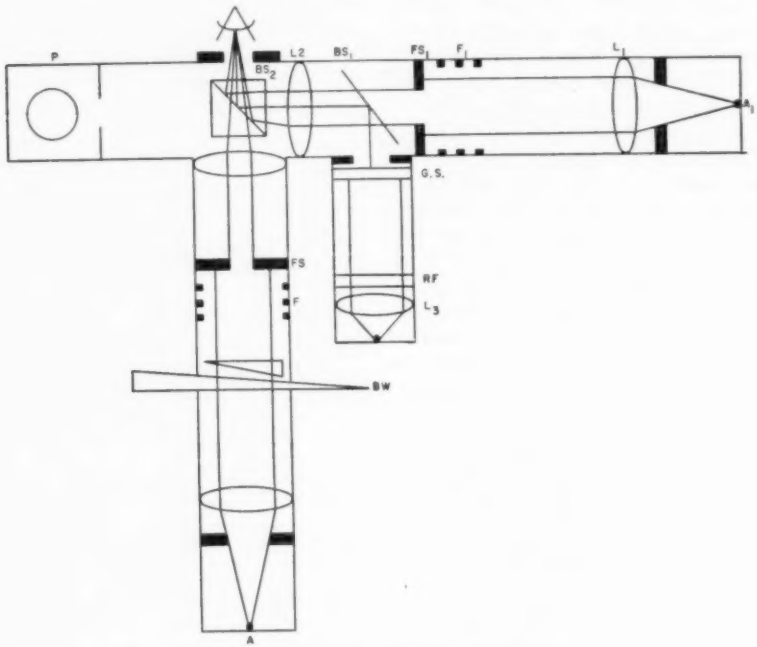

FIG. 1. Schematic of the optical apparatus.

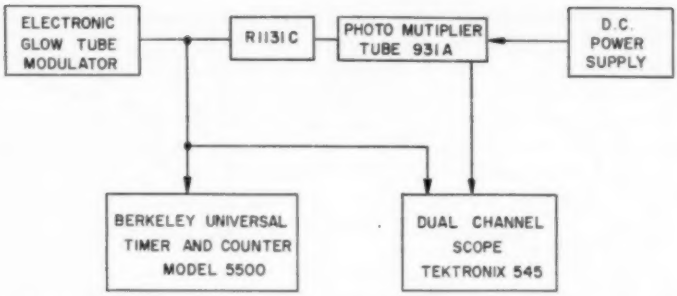
ELECTRONIC GLOW TUBE MODULATOR
R1131C
PHOTO MUTIPLIER TUBE 931A
D.C. POWER SUPPLY
BERKELEY UNIVERSAL TIMER AND COUNTER MODEL 5500
DUAL CHANNEL SCOPE TEKTRONIX 545

FIG. 2. Block diagram of the calibration and monitoring instrumentation.

displayed on the second channel of the Tektronix scope. Since the sweep speed was calibrated, photographing the scope face provided a permanent record of the calibration. In addition, since the voltage drop across the glow modulator tube and the photomultiplier tube output were both displayed on the same scope, their wave forms could be compared. With the exception of a small spike, 6 μsec. in duration, which appeared in both the voltage drop across the glow tube and in the response of photomultiplier, the wave form appeared square.

Prior to each experimental session, all electronic equipment was turned on and allowed to warm up for at least 15 min. After warmup, a calibration check was made on the wave form, duration, and frequency of the test flashes, as well as the light output of the glow modulator tubes. Although the duration of the test flashes was held constant at 1 msec., nine flash frequencies, varying from 1 to 20 flashes/sec, were used in this experiment.

Procedure.—To begin an experimental session *S* would enter a completely darkened cubicle and dark adapted for 15 min. Both *S*s in this experiment were 31-yr-old males having emmetropic vision and experienced as *S*s in visual studies. Dark adaptation was followed by having *S* light adapt to the background luminance for 1 min. The *S*'s position with respect to the eyepiece, which was extended through the wall into the cubicle, was maintained with the aid of a bite board. The background field subtended a visual angle of 14°30′ with the test field subtending 1° of visual angle and appearing in the center of the centrally located fixation field. After the 1 min. of light adaptation, *E* would alert *S* with a buzzer and during the next few seconds present the test stimulus. The *S*'s task was to give one signal with a buzzer when he detected one or more flashes, and two signals with the buzzer when no flashes were detected.

For a given experimental session nine flash frequencies were presented in a randomized order holding flash luminance and background luminance constant. Each frequency was presented 20 times during each session with approximately 20 sec. elapsing between presentations. Three different background luminances and five flash luminances were used, making a total of 15 sessions for each *S*. Five-minute breaks were taken every 15 min., or more frequently at *S*'s request, making an experimental session last about 80 min.

RESULTS

Figure 3 shows the percentage of positive responses as a function of flash frequency for each *S* for three different background luminances, and five flash luminances.[3] Since stimulus presentation was limited to 1 sec., flash frequency also represents the total number of flashes in each test presentation. Each curve represents the data of 1 *S* for one flash luminance at one background luminance. The smooth curves are visual fits with

[3] Complete data for Subjects E and J have been deposited with the American Documentation Institute. Order Document No. 7036 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

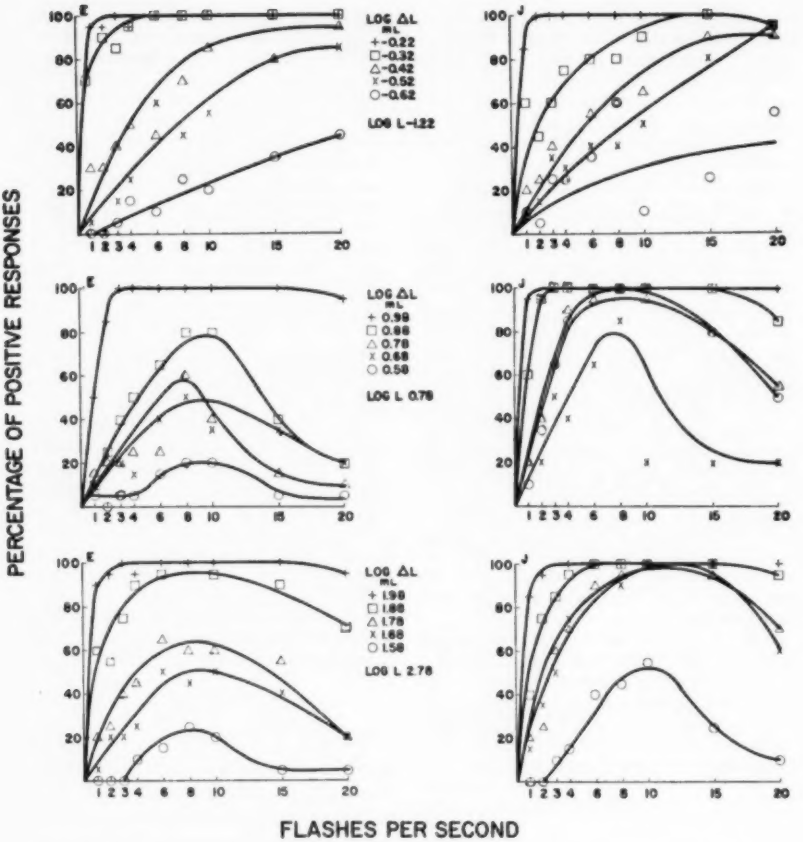

FIG. 3. The percentage of positive responses for 20 presentations as a function of flash frequency for 2 *Ss*, five flash luminances, and three background luminances.

the rise part assuming a probability expectation.

Instead of a smooth regular increase in detection probability with increased flash luminance, an irregular increase occurs with an occasional reversal. These irregularities are probably due to the experimental procedure. Since the primary independent variable was flash frequency, the test flash luminances were selected by pilot work before the regular experiment. Once selected, the flash luminance was held constant for each experimental session. Thus, the observed irregularities are within the session to session variability that occurs in gathering brightness discrimination data. In fact, Mueller (1951) has shown that, using the same apparatus and same *S*, a session to session variability of 0.3 log units of test luminance can be expected.

Under every condition, both *Ss* showed a systematic increase in detection probability with increasing flash

luminance. For the lowest background luminance, there is also a systematic increase in detection probability with increased flash frequency. However, for the higher two background luminances there is a systematic increase in detection probability up to a flash frequency of approximately 10 flashes/sec. At flash frequencies of 15 and 20 flashes/sec, a systematic decrease occurs in detection probability. This decrease occurs even though there is an increase in the number of flashes delivered, which represents an increase in the amount of energy delivered and should therefore represent an increase in the opportunity for detection to occur.

With the intermittent stimulation consisting of a series of constant duration flashes of equal luminance, it is possible to assume that each flash is an independent opportunity for detection and determine the probability of detection on a chance basis. If one assumes that other parameters which might influence the probability are constant, the formulation could take the following form:

$$P = 1 - (1 - p)^n$$

P is the probability of seeing at least one flash.

p is the probability of seeing a single flash in a single flash presentation.

n is the number of flashes presented.

One test of this formulation would be to take the empirical P value for any given n value and compute the other values. However, a limited number of observations were gathered under each condition, and there would be a considerable variation in the expected curves depending upon which specific n was selected.

An alternative test could be obtained by plotting the log of $1 - P$ as a function of n. This should result in a straight line with a slope of log $1 - p$. Figure 4 shows the data of S E for the lowest background luminance plotted in this manner. Since the log of $1 - P$ becomes indeterminate as a probability of 1.0 is reached, only four curves are shown with a fit obtained by a least squares method.

The data plotted in Fig. 4 do not show any systematic deviation from the linear fit. However, it is evident from Fig. 3 that a systematic deviation would occur with this type probability formulation at flash frequencies of 15 and 20 flashes/sec at the higher background luminances.

To compare more readily the curves shown in Fig. 3 with this probability formulation, Fig. 5 was prepared by taking the values of $1 - p$ obtained in Fig. 4 and computing theoretical curves for P. In this way, all the

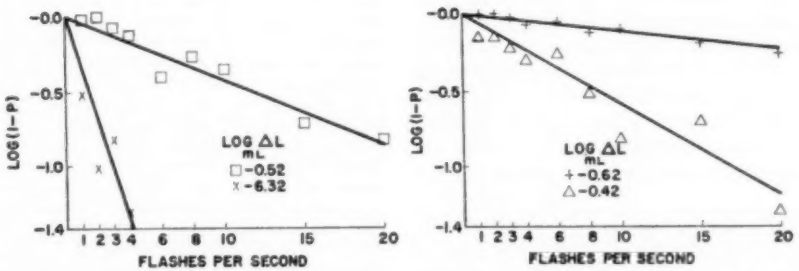

FIG. 4. Curves based upon data of Subject E for a background luminance of log −1.22 mL., with log $1 - P$ plotted as a function of n, with flash luminance as a parameter.

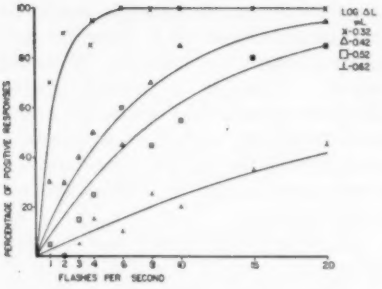

FIG. 5. Curves based upon the data of Subject E for a background luminance of log −1.22 mL., with the percentage of positive responses plotted as a function of flash frequency and flash luminance as the parameter. (The curves show a least squares solution from the data plotted in Fig. 4.)

empirical data contributed to the expected curve. Figure 5 shows the expected and observed values of P plotted as a function of n with flash luminance the parameter. Both Fig. 4 and 5 tend to support the notion that the increase in detection probability with increasing flash frequency is due primarily to the increased number of chances available to make such a detection.

The data of this experiment do not seem adequate to reject the probability formulation except at the higher luminance levels for frequencies of 15 and 20 flashes/sec. Since these decreased probabilities reject the independence assumption, two further assumptions, that preceding flashes had a cumulative effect, either positive or negative, upon succeeding flashes, were evaluated. In both cases, the only effect was to systematically change the rate at which the detection probabilities increased.[4]

[4] Apparently, increasing the number of flashes has two effects: (*a*) a positive and necessary enhancement of the probability of detection due to the extra opportunity for detection, and (*b*) an effect upon flash detection which is related to the spacing between flashes. If the spacing between flashes is great enough, the flashes represent independent opportunities for detection. As the spacing between flashes decreases, a time between flashes interval is reached where a decrement in detection probability occurs. With an additional decrease in the spacing between flashes, a time interval is reached where the energy of two flashes summate and detection probability is again enhanced. This latter effect, the time interval over which the eye will integrate energy, is dependent upon adaptation level (Graham & Kemp, 1938), and the decreased probabilities, which occur at the higher background luminances, might be hidden by the integration effect at the lowest background luminance. Thus, the rejection of the above probability formulations in this experiment is probably due to both the variable spacing effect and the effect of background luminance.

DISCUSSION

Within critical duration, Davy (1952) found that the breaking up of a light stimulus into multiple flashes had no effect upon threshold. Beyond critical duration, threshold energy rises, until, at a separation of 0.5 sec., it is the same for each of two flashes as for one flash and *S*s report two flashes.

If the time interval is great enough, Clark (1958) has shown that a simple probability formulation can account for the increase in detection probability for two flashes over a single flash. The present experiment indicates that the formulation can probably be extended to a somewhat higher number of flashes, particularly if the background luminance is low or the time between flashes is extended.

Perhaps the decreased detection probability is related to the response of a single cell to two light flashes reported by Hartline, Wagner, and MacNichol (1952). This work demonstrated that a flash of light too weak to elicit a response had some excitatory effect that persisted for more than 2 sec. The excitatory effect was demonstrated by using a second flash to test the sensitivity of the receptor at various times after exposure to a subliminal flash. Following a period of enhanced sensitivity to the second flash

there was a period of diminished sensitivity followed by a return to a normal excitatory state.

Similarly, in a train of flashes, the same relative time sensitivities might exist. When flash frequency is low, the receptor unit has time to return to its normal excitatory state and each flash represents an independent opportunity for detection. As flash frequency is increased, a time interval between flashes is reached corresponding to the period of diminished sensitivity. Thus, if the flashes in a train are timed so that each flash after the first arrives during this period of diminished sensitivity, then one might expect a decrease in the probability of detection. With a further increase in flash frequency, a time period is reached corresponding to enhanced sensitivity and one flash contributes to the detection of a following flash.

Within critical duration, Long (1951) showed the wave form of energy distribution to be independent of its effect, and Davy (1952) indicated that the number of flashes times luminance equals a constant. These experiments demonstrate that a frequency or time interval is reached where integration of energy must occur. Further increases in frequency of constant duration pulses simply present more energy to the eye and lead to an apparent brightening of the test field.[5]

If a period of diminished sensitivity accounts for the observed decreased detection, what accounts for the failure to observe this effect at the lowest background brightness? The data of Graham and Kemp (1938) demonstrate that critical duration increases as background luminance decreases. This finding indicates that the lower the background luminance, the longer the time interval during which two flashes would be integrated. In fact, their data indicate summation would occur above 14 cycles/sec for the lowest background luminance used in this experiment and above 20 and 30 cycles/sec for the two higher luminance levels. Thus, one might expect the diminished sensitivity, which exists for flash frequencies of 15 and 20 flashes/sec at the higher background luminances, to be hidden, replaced, or counteracted by the temporal summation effect at the lowest background luminance.

[5] Although only a few observations were gathered after conclusion of the regular experimental session, *Ss* of this experiment indicate that this predicted effect is the observed effect.

Summary

The probability of detection of constant duration intermittent flashes was studied as a function of flash frequency with flash luminance and background luminance as parameters. The *Ss* were adapted to a 14°30′ field of constant luminance and presented with a 1° test flash train in the center of the field. The duration of the stimulus train of flashes was 1 sec. Twenty observations were gathered on 2 practiced *Ss* for each of nine frequencies ranging from 1 to 20 flashes/sec.

The results indicated: (*a*) increasing flash luminance increases the probability of detection for every condition; (*b*) for the condition of low background luminance, probability of detection increases as flash frequency increases; and (*c*) for the higher background luminances, there is a systematic increase in detection up to flash frequencies of approximately 10 flashes/sec, followed by a systematic decrease in probability of detection for frequencies of 15 and 20 flashes/sec.

The increase in the number of chances for flash detection with increasing flash frequency may account for the initial increase in probability of detection that occurs for all luminance values. The decreases in detection probability for the higher flash frequencies at the higher background luminance levels can possibly be attributed to a period of diminished sensitivity that is dependent upon the time interval between flashes. The failure to observe this effect at the lowest background luminance can possibly be attributed to a difference in critical duration.

References

Bartley, S. H. *Vision.* New York: Van Nostrand, 1941.

Clark, W. C. Visual detection threshold as a function of pulse duration and separation. Unpublished doctoral dissertation, University of Michigan, 1958.

Davy, E. The intensity-time relation for multiple flashes of light in the peripheral

retinal. *J. Opt. Soc. Amer.*, 1952, **42,** 937–941.

GRAHAM, C. H., & KEMP, E. H. Brightness discrimination as a function of the duration of the increment in intensity. *J. gen. Physiol.*, 1938, **21,** 635–650.

HAMBURGER, F., JR., & KING, E. J. Recording photometer and its use in studies of cathode-ray screen displays. *J. Opt. Soc. Amer.*, 1948, **38,** 875–879.

HARTLINE, H. K., WAGNER, H. G., & MACNICHOL, E. F., JR. The peripheral origin of nervous activity in the visual system. *Symp. Quant. Biol.*, 1952, **17,** 125–141.

LONG, G. E. The effect of duration of onset and cessation of light flash on the intensity-time relation in the peripheral retina. *J. Opt. Soc. Amer.*, 1951, **41,** 743–747.

MUELLER, C. G. Frequency of seeing functions for intensity discriminations at various levels of adapting intensity. *J. gen. Physiol.*, 1951, **34,** 463–474.

SCHUIL, A. E. The fixed light equivalent of flashing lights having constant duration. *Trans. Illum. Engng. Soc., Lond.*, 1940, **5,** 117.

(Received January 3, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 361–369

THE TEMPORAL COURSE OF PROACTIVE INHIBITION FOR SERIAL LISTS[1]

LEO POSTMAN

University of California

Application of interference theory to the analysis of long-term retention has brought into focus the importance of proactive inhibition (PI) as a mechanism of forgetting. As Underwood (1957) has shown, classical studies of retention using practiced *Ss* in counterbalanced designs greatly overestimated the amount of forgetting because of substantial amounts of PI accumulated during the learning of successive lists. Retention of single lists by naive *Ss* is considerably higher than had been found with practiced *Ss* in earlier investigations (Postman, 1961; Postman & Rau, 1957; Underwood & Postman, 1960; Underwood & Richardson, 1956). The retention losses observed with naive *Ss* may reasonably be attributed to retroactive inhibition (RI) and PI from verbal habits practiced outside the laboratory. More of these habits are likely to have been acquired prior to the experiment than during the retention interval. Hence extraexperimental interferences should be largely proactive rather than retroactive.

The conditions and characteristics of extraexperimental interference are assumed to be the same as those observed in formal studies of RI and PI. Specifically, the acquisition of the experimental list is believed to involve the unlearning or extinction of pre-experimental associations which *S* brings to the laboratory. The extinguished associations recover during the retention interval and compete with the prescribed responses at the time of recall. The amounts of extinction and recovery depend, of course, on the number and strength of the pre-experimental associations in competition with the prescribed responses. Thus, there should be more competition from extraexperimental sources for words of high than low meaningfulness. The patterns of errors in learning and recall obtained with materials of high and low meaningfulness are consistent with this expectation (Postman, 1961; Underwood & Postman, 1960).

A central assumption of the present analysis of forgetting is that PI increases as a function of time. This assumption may be considered correct by definition. In experiments on interserial interference, PI is obtained even if the learning of the second list is carried to a point at which associative inhibition is no longer present or is superseded by associative facilitation. The necessary condition for obtaining PI is that the retention interval be longer than the intertrial interval in the acquisition of the second list. If the process responsible for the interference is continuous, PI must increase as a function of time (Underwood, 1945, 1948a). It is important to ask, however, whether the rate of increase in PI makes it reasonable to attribute long-term forgetting largely to proactive interference from extraexperimental sources.

Experimental evidence on the temporal course of PI for verbal habits is not extensive. One study explicitly concerned with this problem is that of Underwood (1949) which used paired adjectives, with practiced *Ss* serving in a counterbalanced design. No increase in PI was found between 20 min. and 75 min. after the end of learning. There was, however, a substantial increase in interlist intrusions, indicating a loss of differentiation as a function of time. In

[1] This research was supported by a grant from the National Science Foundation.

another experiment using paired adjectives, Underwood (1948a) compared RI and PI after 5 and 48 hr. After 5 hr. RI was significantly greater than PI. This difference was no longer present after 48 hr.; retention of the second list had declined whereas there was a slight increase in the recall of the first list. Since PI is a function of the degree of prior learning, the difference in the rate of forgetting for the two lists suggests that there should be long-term increases in PI. The results of experiments using the method of modified free recall (MFR) lead to the same prediction since they show long-term increases in the frequency of responses from the first list along with a steady decline in responses from the second list (Briggs, 1954; Briggs, Thompson, & Brogden, 1954; Underwood, 1948b).

This experiment investigates the temporal course of PI for serial lists with naive *Ss*. As in some of the recent studies of extraexperimental interference (Postman, 1961; Underwood & Postman, 1960), words of high (HF) and low frequency (LF) of usage were used. It was expected that (*a*) PI at recall would increase significantly during a 1-week retention interval, and (*b*) the loss of differentiation, as measured by interlist intrusions, would be greater with materials of high than low frequency. The latter expectation was based on the assumption that there would be greater overlap in the associative context of successive lists with materials of high frequency.

Method

Design.—The three independent variables were Condition of Practice (Experimental vs. Control), Word Frequency (High vs. Low), and Retention Interval (30 sec., 30 min., and 7 days). The Experimental groups learned three successive lists and were tested for recall of the third. The Control groups learned and recalled a single list. The results of the 30-sec. test were used to assess associative strength at the end of learning, in view of expected differences in rate of acquisition as a function of word frequency and interlist transfer. With two conditions of practice, two levels of word frequency, and three retention intervals, the total design comprised 12 independent groups.

Lists.—The learning materials were lists of 12 one-syllable nouns. There were three HF lists and three LF lists. For the HF words the number of occurrences in the "L" count of Thorndike and Lorge (1944) was between 1000 and 3300 in 4.5 million; for the LF words the number of occurrences ranged between 1 and 3. There was no repetition of first letters in any of the lists. Six different serial orders were used equally often with each of the lists. The orders were so constructed that no word appeared more than once in any serial position and that no forward sequence of two words was used more than once.

There were no synonyms among the words of a given frequency. At each frequency level, the three lists had 10 first letters in common; 2 of the first letters in each list were not repeated in that position in the other lists. For the Experimental groups, the six possible orders of the three lists were used equally often. For the Control groups each of the single lists was used equally often. The lists are presented in Table 1.

Procedure.—The lists were presented on a Hull-type memory drum at a 2-sec. rate, with an intertrial interval of 6 sec., and learned by the anticipation method to a criterion of one perfect trial. In the experimental condition, successive lists were separated by an interval of 1 min. Retention was tested by relearning for five trials or to a criterion of one perfect trial, whichever took the longer. For the groups tested after 30 min., the retention

TABLE 1

Lists Used in the Experiment

High Frequency			Low Frequency		
List 1	List 2	List 3	List 1	List 2	List 3
BED	BOY	ARM	BOLE	BARD	BRIG
CHILD	CHAIR	BILL	CARP	CRYPT	CAW
DRESS	DROP	CAR	FINCH	FLUME	DRAM
FLOOR	EYE	DOOR	GIG	GUILE	FOUNT
JOB	FRONT	FIGHT	JAG	LOUT	GROAT
KISS	LINE	MIND	LEEK	MEIN	IRE
MILE	MONTH	NEED	MIDGE	PUG	LODE
NAME	NIGHT	PRICE	PRAWN	QUOIT	MALL
PART	POST	SIDE	SKEIN	SKIFF	PRISM
SENSE	STATE	TRAIN	TUSK	TALC	SLAG
TOUCH	TOWN	VOICE	VANE	WART	TITHE
WALL	WIFE	WORLD	WOMB	YAP	WENCH

Note.—Six different random orders were used with each list.

interval was filled with two tasks—mathematical problems (20 min.) and rating of pictures (10 min.).

Subjects.—There were 12 groups of 12 *Ss* each. The *Ss* were assigned to conditions in blocks of 12, with 1 *S* per block from each of the groups of the experimental design. The running order within each block was determined by a table of random numbers, as was the assignment of lists and serial orders within each condition. All *Ss* were undergraduate students who had English as their native language and were naive to rote-learning experiments. No *Ss* had to be rejected because of failure to learn.

Results

First-list learning.—For List 1—the first of the two prior lists of the Experimental groups and the test list of the Control groups—there were no significant differences among the groups learning material of a given frequency. The results of the 72 *Ss* learning each type of list have, therefore, been combined. The mean number of trials to criterion for the HF lists was 16.83 (SD = 6.07), and 21.14 (SD = 8.02) for the LF lists. The difference in speed of learning is significant beyond the .001 level (t = 3.59, df = 142).

The mean number of misplaced responses per trial was 1.46 (SD=.63) for the HF lists and 1.22 (SD = .51) for the LF lists. This difference is highly significant (t = 8.00, df = 142, $P < .001$). Misplaced responses account for all but a few of the overt errors during acquisition. The mean percentages that misplaced responses were of all overt errors are 98.4 and 95.1 for the HF and LF lists, respectively. The difference in favor of the HF groups is maintained when the misplaced responses are expressed as percentages of opportunities (total occurrences minus correct responses). The mean percentage is 28.4 (SD = 13.9) for the HF lists, and 22.4 (SD = 9.7) for the LF lists. The difference in percentages (following arc-sine transformation) is also significant (t = 2.65, $P < .01$). As in an earlier study using two-syllable nouns from the same frequency ranges, speed of learning and error rate are directly related (Postman, 1961).

Transfer effects in original learning.—The mean numbers of trials to criterion for the three successive lists learned by the Experimental groups are shown in Table 2. For both kinds of materials, there is a substantial increase in speed of learning on List 2 and little further change on List 3. The relative amounts of improvement are comparable for the HF and LF lists. The effects of order of learning are significant (F=112.62, df=2/140,

TABLE 2

Trials to Criterion and Results of Stage Analyses for Successive Lists Learned by the Experimental Groups

List	Word Frequency			
	High		Low	
	Mean	*SD*	Mean	*SD*
	Trials to Criterion			
1	17.25	6.08	21.31	6.95
2	10.06	3.73	11.67	4.70
3	8.69	3.65	11.78	5.48
	Trials for Response Learning[a]			
1	4.24	1.72	5.24	1.36
2	2.97	.82	3.99	1.14
3	2.89	.96	3.89	.85
	Trials for Associative Learning[b]			
1	1.39	1.01	1.69	1.62
2	.40	.47	.34	.38
3	.29	.34	.25	.35

[a] Mean trial on which response was given anywhere in the list.
[b] Mean number of trials between first occurrence and first correct placement of response.

$P < .001$) but do not interact with frequency ($F = 1.67$).

Since there was substantial "learning to learn," the test was mastered considerably faster by the Experimental groups than the Control groups. Differences in learning are likely to influence retention, and it becomes important to inquire what characteristics of performance were responsible for the positive transfer effects. We considered first the effects of practice on the two stages into which the period of acquisition can be divided analytically—the response-learning stage and the associative stage (Underwood & Schulz, 1960, pp. 92–94). The duration of the response-learning stage is estimated by the mean trial on which the correct responses are given anywhere in the list. The mean number of trials between the first occurrence and the first correct placement of the responses measures the length of the associative stage. Table 2 shows the results of stage analyses for the three successive lists learned by the Experimental groups. The duration of the response-learning stage declines sharply for both kinds of materials between List 1 and List 2 but does not drop further on List 3. The variation among lists is significant ($F = 44.44$, $df = 2/140$, $P < .001$) but order of learning does not interact with frequency ($F < 1$). Throughout practice, the response-learning stage remains shorter for the HF than the LF lists. The over-all effect of frequency is significant beyond the .001 level ($F = 21.90$, $df = 1/70$).

The changes in the duration of the associative stage parallel those for the response-learning stage. Again there is a clear increase in speed between Lists 1 and 2 and no further change on List 3. Order of learning is significant ($F = 17.84$, $df = 2/140$, $P < .001$) whereas the interaction of order with frequency is not ($F < 1$). In contrast to the results obtained for the response-learning stage, the differences between the two kinds of materials are small and inconsistent. The over-all effect of frequency is not significant ($F < 1$). It may be noted that the same conclusions concerning the effects of frequency on the two stages of learning are reached when the performance of all *S*s on List 1 is considered. The difference between the HF and LF conditions is significant for the response-learning stage ($t = 4.52$, $df = 142$, $P < .001$), but not for the associative stage ($t = 1.55$).

A second systematic effect of practice is the reduction in the relative frequency of misplaced responses. Table 3 presents for Lists 1–3 the mean numbers of misplaced responses per trial as well as the mean percentages of such errors based on the number of opportunities. The mean number of errors per trial declines from list to list whereas there is a clear drop in the percentages only on List 3. A decline in the number of errors per trial may merely be complementary to an increase in correct responses, and the percentages probably provide the clearest index of *S*s' disposition to give misplaced responses. In any event the rate of misplaced responses is clearly lower for List 3 than for List 1. The change in both indices is significant beyond the .01 level ($F = 5.36$ for responses per trial and 7.72 for percentages, $df = 2/70$ in both cases). In neither case is there a significant interaction between order and frequency ($F = .06$ and 1.78, respectively).

Interlist intrusions during learning.—The number of interlist intrusions during the acquisition of Lists

TABLE 3

MEAN NUMBERS PER TRIAL AND PERCENTAGES OF MISPLACED RESPONSES IN ACQUISITION OF SUCCESSIVE LISTS BY EXPERIMENTAL GROUPS

List	Word Frequency					
	High			Low		
	Mean	*SD*	%	Mean	*SD*	%
1	1.40	.57	26.8	1.20	.52	22.0
2	1.00	.45	26.2	.91	.62	19.7
3	.84	.46	23.0	.68	.47	15.3

2 and 3 was small. Under the HF condition, a total of 5 intrusions was contributed by 3 *S*s during the learning of List 2, and a total of 10 intrusions by 7 *S*s during the learning of List 3. All of the latter came from List 2. Under the LF condition, there were 4 intrusions obtained from 3 *S*s during the learning of List 2, and 9 intrusions from 7 *S*s during the learning of List 3. Of these 9 intrusions, 6 came from List 1 and 3 from List 2. There is a suggestion here that the extinction of List 1 responses during acquisition of List 2 was less effective with materials of low than high frequency.

Recall.—The mean numbers of items recalled by the various groups are shown in Fig. 1. In view of the differences in rate of acquisition of the test list as a function of both frequency and interlist transfer, the results of the 30-sec. tests are of interest. Under the control condition, the postcriterial drop is, as expected, greater for the LF list than the HF list. Under the experimental condition, however, the reverse is true in spite of the fact that the HF list was learned faster than the LF list. This reversal reflects the fact that with HF materials the drop is greater for the Experimental than the Control group whereas with LF materials the difference is in the opposite direction. For the results of the 30-sec. tests, the interaction of condition with frequency approaches, but does not reach, significance ($F = 3.84$, whereas 4.08 is required at the .05 level with 1/44 *df*). The pattern of differences suggests that proactive effects may develop extremely rapidly with HF materials and may influence even a retention test immediately after the end of learning. The fact that there were no interlist intrusions on the 30-sec. tests does not contradict this interpretation since differentiation of the lists should be high at this point. In any event, it is important to note that the usual inverse relationship between speed of learning and postcriterial drop does not necessarily hold true under conditions of transfer.

We turn now to the results of the delayed tests. Under the control condition the rate of forgetting for the two kinds of materials is comparable. During the 1-week interval (30 sec.–7 days) the mean retention

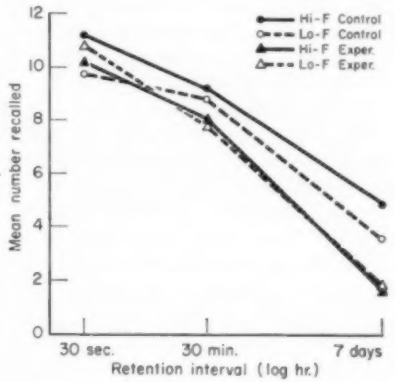

FIG. 1. Mean numbers of items correctly recalled by the Experimental and Control groups as a function of retention interval.

loss is 6.25 for the HF lists, and 6.17 for the LF lists. The Experimental groups recall less than the Control groups, and this difference increases for both types of lists between 30 min. and 7 days. Following a Freeman-Tukey square-root transformation, the recall scores were subjected to an analysis of variance. The differences among retention intervals are significant beyond the .001 level ($F = 137.21$, $df = 1/132$) as is the difference between Conditions ($F = 13.55$, $df = 1/132$). The interaction, Time × Conditions, is also significant ($F = 7.25$, $df = 2/132$, $P < .01$) and reflects the increasing difference between Experimental and Control groups as a function of time. In order to determine whether there was a reliable increase in PI between 30 min. and 7 days, the interaction variance was partitioned into two components with 1 *df* each—that associated with the differences between the 30-sec. test and the combined delayed tests on the one hand, and that associated with the differences between the two delayed tests on the other. The two F ratios of 7.73 and 6.78 are significant beyond the .01 and .02 levels, respectively. These tests confirm the conclusion that there were progressive increases in the amount of PI.

The forgetting curves shown in Fig. 1 suggest that the divergence between the Experimental and Control groups is more pronounced for the HF than the LF lists. However, neither frequency nor any of the interactions involving frequency is a significant source of variance. Thus, the rate of increase in PI is not reliably different for the two kinds of materials.

Misplaced responses at recall.—The number of different anticipatory errors was used to measure the tendency to misplace correct responses at recall. Each such error represents a response which is available to *S* before re-exposure to the item but whose serial position has been forgotten. As Fig. 2 shows, the degree of recovery of these errors corresponds to the rate of misplaced responses during acquisition of the test list. Thus, the errors recover more rapidly for the Control groups than the Experimental groups, and for the HF lists than the LF lists. In view of the large number of zero scores, the following procedure was used to evaluate the significance of the temporal trends. Each *S* in the 30-min. and 7-day groups was matched with the *S* in the appropriate 30-sec. group in the same block. The difference in the number of errors was determined for each pair, and the distributions of difference scores were subjected to an analysis of variance. The scores of the 2 *S*s matched to the same 30-sec. *S* were treated as repeated measurements. The variations as a function of frequency and condition are significant beyond the .05 level ($F = 4.25$, $df = 1/44$ in both cases since the mean differences were

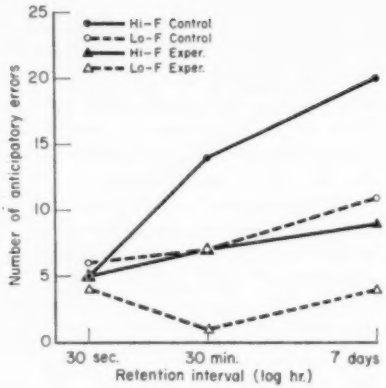

FIG. 2. Numbers of different anticipatory errors at recall.

the same for the two effects). The difference between 30 min. and 7 days is not significant ($F = 2.55$, $df = 1/44$), and there are no significant interactions. The relatively small increases in anticipatory errors between 30 min. and 7 days probably reflect reduced availability of responses from within the list. The increases are larger when total numbers of errors rather than numbers of different errors are considered since *Ss* with low recall scores tend to repeat the same errors several times in succession.

Interlist intrusions at recall.—Table 4 shows the frequencies of interlist intrusions at recall and the numbers of *Ss* contributing them. There were no intrusions on the 30-sec. tests. The total numbers of intrusions on the delayed tests are small but there are two consistent trends. First, there is an increase in intrusions as a function of time which may be attributed to a loss of list differentiation. Second, more intrusions are given during the recall of the HF lists than the LF lists, indicating that the loss of differentiation is greater for the former material. Finally, most of the intrusions are from the second of the two prior lists. This finding supports Underwood's (1945) conclusion that contiguity during learning is conducive to interlist confusion.

TABLE 4

NUMBERS OF INTERLIST INTRUSIONS AT RECALL

Word Frequency	Retention Interval			
	30 Min.		7 Days	
	Total	N^a	Total	N^a
High	4(4)	2	8(6)	5
Low	—	—	3(3)	2

Note.—Numbers from second prior list are in parentheses.
[a] Number of *Ss* giving intrusions.

TABLE 5

MEAN NUMBERS OF CORRECT ANTICIPATIONS IN FIVE TRIALS OF RELEARNING

Retention Interval	Cond.	Word Frequency			
		High		Low	
		Mean	*SD*	Mean	*SD*
30 sec.	Exp.	53.42	4.25	53.83	6.03
	Cont.	55.25	2.98	53.33	5.87
30 min.	Exp.	51.75	6.18	47.50	9.84
	Cont.	51.83	5.06	50.42	7.11
7 days	Exp.	33.25	4.99	33.17	10.08
	Cont.	41.08	8.69	39.92	7.49

Relearning.—The mean numbers of correct responses in five trials of relearning are shown in Table 5. The differences between the Experimental and Control groups are quite small on the 30-min. tests but remain substantial on the 7-day tests. The interaction of Time with Conditions is significant ($F = 3.80$, $df = 2/132$, $.02 < P < .05$, after square-root transformation). Thus, the differences in PI as a function of time are not limited to the recall trial. This conclusion is confirmed by the fact that the interaction of Time with Conditions remains significant when the scores on Trial 2 of relearning are analyzed separately ($F = 6.35$, $df = 2/132$, $P < .01$). However, the dissipation of interference is sufficiently rapid to eliminate significant PI when retention is measured by trials to criterion in relearning.

DISCUSSION

The experiment yields clear evidence for substantial increases in PI as a function of time. These results were obtained with naive *Ss* for whom the only controlled source of interference consisted of two lists learned immediately prior to the test list. In the light of these

findings the assumption that PI from extraexperimental sources increases progressively in time remains reasonable. According to an interference interpretation, both the Experimental and Control groups in the present study were subject to PI, and possibly RI, from extra-experimental sources. For the Experimental groups, additional interference came from the lists learned prior to the test list. The prior lists added substantially to the total effective interference at recall since they were associated with the same experimental situation as the test list. Thus, retention decreased progressively for both the Experimental and Control groups but significantly more so for the former.

In assessing the amounts of PI observed in the present study, it is important to consider the differences between the Experimental and Control groups in the acquisition of the test list. As a result of practice on the prior lists, both the response-learning and associative stages of acquisition were shortened for the Experimental groups, the probability of misplaced responses was reduced, and the speed of mastery to criterion was increased. These effects of "learning-to-learn" should produce greater associative strength at criterion under the experimental than the control condition. The results of the 30-sec. tests cannot take full account of all the differences in learning. For example, since prior learning reduced the rate of misplaced responses, there were fewer such errors which could recover for the Experimental than the Control groups. Such differences in error tendencies first become apparent on the delayed tests of retention (Fig. 2). Thus, large amounts of PI developed during the retention interval in spite of the fact that the conditions of acquisition probably were more favorable to retention under the experimental condition.

There were no significant differences between the amounts of PI obtained with the HF and LF lists. This finding parallels the results obtained with single lists: there was no difference in the rate of forgetting for the two kinds of materials either under the control condition of the present experiment or in an earlier study investigating the retention of serial lists (Postman, 1961). There is a strong indication that under the experimental condition the loss of list differentiation was, as expected, greater with the HF than the LF materials. Although interlist intrusions were given by only part of the *S*s, they were consistently more numerous in the recall of the HF lists.

As in other studies of interserial interference, the number of interlist intrusions at recall was small and accounts for only a fraction of the total retention decrements under the experimental condition. The number of intrusions increases as a function of time but at a much slower rate than total PI. On the assumption that the differentiation of lists learned to criterion is lost slowly and is reinstated readily during the recall trial, at least some of the retention loss can be attributed to implicit intrusions which were rejected by *S*. The assumption that differentiation remained high is indirectly supported by the fact that importations from outside the list were extremely rare under any of the conditions of the experiment. Thus, there was highly effective discrimination between the experimental lists and extra-experimental responses as well as among the successive lists.

Summary

This experiment investigates the temporal course of PI for serial lists. Lists composed of either high-frequency words or low-frequency words were used. The Experimental groups learned three successive lists and were tested for recall of the third. The Control groups learned and recalled a single list. Retention was tested by relearning after intervals of 30 sec., 30 min., and 7 days. There were significant increases in the amount of PI as a function of time for both kinds of materials. There were no significant differences as a function of word frequency, either in the retention of single lists or in the amount and rate of increase in PI. Interlist intrusions increased as a function of time and were more numerous in the recall of the high-frequency than the low-frequency lists. The fact that

PI increases substantially over time is consistent with the hypothesis that the growth of proactive interferences from extraexperimental sources is responsible for much of long-term forgetting.

References

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47,** 285–293.

BRIGGS, G. E., THOMPSON, R. F., & BROGDEN, W. J. Retention functions in reproductive inhibition. *J. exp. Psychol.*, 1954, **48,** 419–428.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61,** 97–110.

POSTMAN, L., & RAU, L. Retention as a function of the method of measurement. *U. Calif. Publ. Psychol.*, 1957, **8,** 271–396.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J. The effect of successive interpolations on retroactive and proactive inhibition. *Psychol. Monogr.*, 1945, **59**(3, Whole No. 273).

UNDERWOOD, B. J. Proactive and retroactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38,** 29–38. (a)

UNDERWOOD, B. J. "Spontaneous" recovery of verbal associations. *J. exp. Psychol.*, 1948, **38,** 429–439. (b)

UNDERWOOD, B. J. Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, 1949, **39,** 24–34.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, **64,** 49–60.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67,** 73–95.

UNDERWOOD, B. J., & RICHARDSON, J. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp. Psychol.*, 1956, **52,** 119–126.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received January 10, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 370–375

TRANSFER OF TRAINING AS A FUNCTION OF TASK DIFFICULTY IN A COMPLEX CONTROL SITUATION [1]

DONALD A. GOLDSTEIN AND JOHN M. NEWTON [2]

Electric Boat Division, General Dynamics Corporation

An important problem in training is that of assessing the effects of initial task difficulty on transfer, i.e., whether training on a more difficult task results in better transfer to a less difficult task or vice versa. Studies attempting to answer this question have produced equivocal results. For example, Ritchie and Michael (1955) have found better transfer from more to less difficult tasks while Lordahl and Archer (1958) have found the reverse. Day (1956), in his review of the area, has called attention to a major problem in generalizing about results: the lack of specificity as to how difficulty was varied.

One area in which it appears possible to specify task difficulty is that of complex control systems, i.e., those systems in which one or more integrations, with corresponding time lags for these integrations, intervene between control action and system reaction. To a large extent, differences in ease of control (task difficulty) reflect both the complexity of the control dynamics involved and the length of the time phase relationship between control and display movements. For longer control lags, *S* must attend more closely to the subtle information (rate and acceleration) than he would have to for shorter control lags, given the same number of integrations. In extreme cases, because of the long delay, *S* may be forced to make compensatory control movements before the effects of his previous movement(s) become apparent on the display. Thus, differences in control lag may reflect differences in ease of control (task difficulty). Differences in control lag, however, are themselves a function of certain physical restraints inherent in the system, such as moment of inertia, system gain, etc., and some of these, in turn, can be identified with certain coefficients in the dynamic equations of motion representing the system. By specifying task difficulty in terms of changes in the values of certain coefficients in the dynamic equations of a complex control system, it should be possible to determine experimentally: (*a*) which resulting system, of those chosen to study, proves to be most difficult to control, and (*b*) which system produces, on the average, the greater positive transfer to all other systems studied.

To investigate these effects, coefficients affecting moment of inertia and speed of system reaction in the dynamic equations of motion of a submarine (a third-order integration

[1] This research was supported by the United States Navy under Contract NOnr 3075(00), monitored by the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research, Washington 25, D. C. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. The authors wish to express their appreciation to Lawrence Hill and George Simoneau of the Human Factors Staff for their aid in the design and construction of the apparatus used in this study.

[2] Now at the University of Omaha, Omaha, Nebraska.

system) were varied, each variable taking two separate sets of values. In combination, this procedure defined four separate training conditions. Through the use of a factorial design all possible transfer combinations among the four conditions were observed.

METHOD

Apparatus.—The control-display system used in this study was the same for all *Ss*. The display consisted of a straight scale indicator, 18 in. long, calibrated from ±150 ft. in 10-ft. increments, with ⅛ in. equal to 10 ft. The *S*'s task was to keep a moving pointer, ⅛ in. long, at a zero reading on the scale. The electrical limits of the system were ±100 ft., with 10 v. equal to 100 ft. The error generator employed a sine wave forcing function to move the pointer within a range of ±50 ft. at the rate of .006 cps with an amplitude of ±2.5 v. The position on the sine wave at the start of the trial was randomized across trials.

The control, a modified joystick, 6¾ in. long and 1 in. diameter, had a maximum deflection angle of ±45° with the stick voltage output at full travel equal to ±75 v. The equivalent plane angle at full travel was ±25°, making the angular deflection of the plane roughly equal to one-half the angular deflection of the control stick.

A Heathkit 15 amplifier computer was utilized with 10 analog amplifiers used for the submarine programs and 4 amplifiers used for the scoring circuits. A Standard Electric timer was used to record time on a target zone of ±10 ft. for the last 100 sec. of each 120-sec. trial. The programs employed were simplified equations of submarine motion in the *Z* axis, with differing constants depending upon the moment of inertia and speed of reaction of the submarine being simulated.

Table 1 shows the simplified equations of motion used in the computer program. These equations are a result of maintaning a constant speed within any trial. Values of the constants are not presented since they are completely artificial.

Subjects.—The *Ss* were 80 undergraduate students, 58 men and 22 women, enrolled at Mitchell College in New London, Connecticut. An additional 5 *Ss* had to be eliminated from the study for failure to follow directions or as a result of equipment breakdown. All *Ss* were volunteers taken from undergraduate courses and they were paid $1.00 per hour for their participation.

TABLE 1

SIMPLIFIED EQUATIONS OF MOTION USED IN THE COMPUTER PROGRAM

Type of Motion	Equation
For pitch:	$\ddot{\theta} = -k_1\theta - k_2(k_{12}\theta - k_3 W) \pm k_4\delta_s$
For normal force:	$\dot{W} = k_5\delta_s + k_6\dot{\theta} - k_7 W$
For depth rate:	$\dot{Z} = k_8\theta - k_9 W$
For depth:	$Z = k_{10}\dot{Z}$

where: θ = displacement in the horizontal attitude
Z = displacement in the vertical attitude
U = axial velocity and is contained within the k values
W = virtual mass
δ_s = displacement of the control stick
$\ddot{}$ = acceleration terms
$\dot{}$ = rate terms

Design.—The factorial design for this study employed four dimensions: moment of inertia trained on, moment of inertia tested on, speed of system reaction trained on, and speed of system reaction tested on. As each of these dimensions had two levels, large and small moment of inertia, fast and slow speed of system reaction, the final analysis resulted in 16 cells ($2 \times 2 \times 2 \times 2$), with five replications within each cell.

Procedure.—The 80 *Ss* were randomly assigned to 1 of 16 groups, with 4 groups training under each of the original conditions. These conditions represented the effects of a small moment of inertia with a fast speed of system reaction (SF), a small moment of inertia with a slow speed of system reaction (SS), a large moment of inertia with a fast speed of system reaction (LF), and a large moment of inertia with a slow speed of system reaction (LS). After training, each of the 4 groups for each original condition was transferred to the same condition or one of the other three conditions.

All *Ss* were tested individually and detailed instructions as to the nature of the task (depth-keeping) as well as information on the behavior of control-display operations were presented to them on tape. The *Ss* were permitted one free trial while the instructions were being read and were allowed to ask questions prior to the start of the experiment.

During training each *S* was given 20 120-sec. trials with a 10-sec. rest between each trial within a 10-trial block. Between each 10-trial block there was a 3-min. rest period. At the completion of the training trials, a 5-min. rest period was given each *S* followed by 10 test trials. The same inter-and intratrial intervals were used for all conditions.

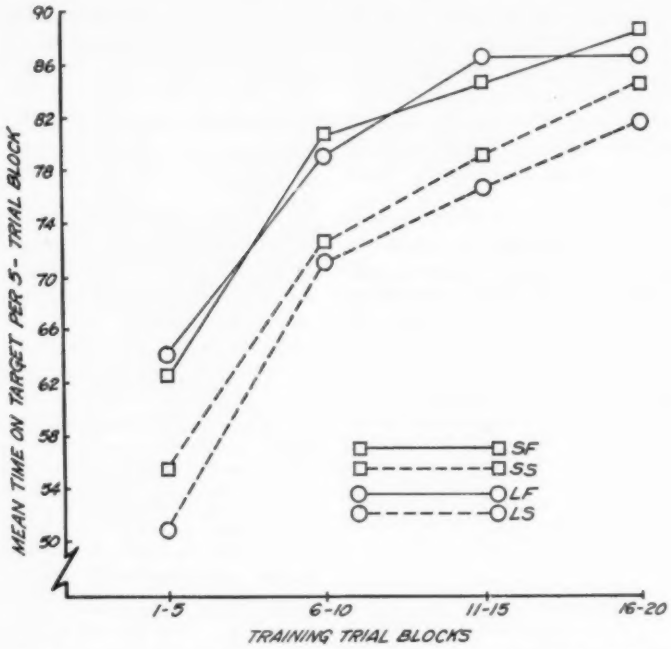

FIG. 1. Mean time on target for Cond. SF, SS, LF, and LS (N = 20 in each group).

The proficiency measure taken was the time on a target zone of ±10 ft. No score was recorded for the first 20 sec. of each trial to allow S time to bring the pointer to the ordered position. Knowledge of results was provided by informing S of his time on target score following each trial.

RESULTS

The dependent-variable units used in the analyses were each S's mean time on target score ($\bar{X}_{TOT}$) per five-trial block. The relation between system dynamics and performance for the training data is presented graphically in Fig. 1. The general trend, clearly shown, was for better performance under Cond. SF and LF. An analysis of variance across trial blocks also was performed on the data.[3] The results showed that the main variable affecting performance was the speed of system reaction. Those groups performing the depth-keeping task with the fast speed of reaction did better than those groups operating with the slow speed ($F = 4.16$, $df = 1/76$, $P < .05$). The main effect "trial blocks" also proved significant ($F = 93.48$, $df = 3/228$, $P < .01$), which indicated a general improvement in performance over trials.

To determine the effects of prior training on transfer, a method for evaluating these effects when Ss are trained under different initial conditions and are performing at different levels at the time of transfer was sought. Two methods employed in studies concerned with a similar problem were: (*a*) to compare per-

[3] Bartlett's test was performed on these and all other data used in subsequent analyses with no test value being significant at the .05 level.

formance on the test condition for one group with performance on the initial condition for the other group (Holland & Henson, 1956; Ritchie & Michael, 1955), and (*b*) to utilize a condition of intermediate difficulty, equally unfamiliar to all groups, for the test condition (Lordahl & Archer, 1958). As neither method appeared appropriate in view of the design employed, a multivariate analysis of covariance was used to equate for differences in terminal performance levels in training prior to the imposition of test condition treatments. This analysis was performed on the first five-trial block of the test trials with the adjusting variable used being the $\bar{X}_{TOT}$ scores for the last five-trial block of training trials. The test of significance of the regression of the dependent variable on the

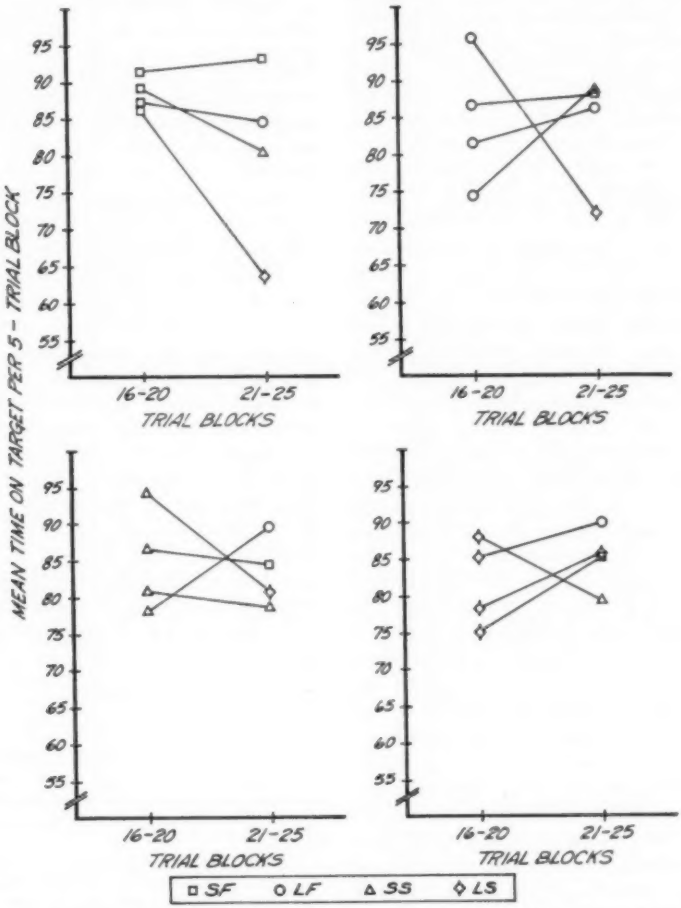

FIG. 2. Mean time on target for all transfer conditions using adjusted sample means ($N = 5$ in each group).

adjusting variable indicated a satisfactory degree of relationship ($F = 46.35$, $df = 1/62$, $P < .01$). This finding supported the use of the covariance design as a means of removing part of the variance representing initial differences among the experimental units.

Using adjusted sample means the data of this analysis are plotted separately for each transfer condition in Fig. 2. The results of the analysis showed that speed of system reaction was the only main variable affecting performance on the test trials. Those groups performing the task with the fast speed of reaction did better than those groups operating with the slow speed ($F = 10.11$, $df = 1/63$, $P < .01$). This result is not surprising. It supports the finding on the training data analysis in that speed of system reaction is still the important variable governing performance.

The first order interaction, speed of system reaction trained on and moment of inertia tested on, also was found to be significant ($F = 7.94$, $df = 1/63$, $P < .01$). In order to focus on the source of the significant F, the 16 samples were regrouped according to the classification of the interaction, using adjusted sample means. The means for each of the four samples making up the cells of the 2×2 matrix displaying the interaction are presented in Table 2.

Inspection of Table 2 indicates that the worst performance occurred for those *S*s previously trained on a fast speed of reaction system and subsequently tested on a large moment of inertia system. The best performance was exhibited by those *S*s trained on a fast speed of reaction system and tested on a small moment of inertia system. However, training on a slow speed of reaction system resulted in better performance on both moment of inertia systems when compared with performance on a large moment of inertia system following training on a fast speed of reaction system.

TABLE 2

MEANS OF SAMPLES REPRESENTING THE INTERACTION, SPEED OF SYSTEM REACTION TRAINED ON (SR_{TR}) × MOMENT OF INERTIA TESTED ON (MI_{TE})

Moment of Inertia Tested On	Speed of System Reaction Tested On	
	Fast	Slow
Small	87.74	81.92
Large	76.40	86.45

Using Lindquist's (1953) "least significant difference" method for comparing means two at a time, two of the above mentioned differences were found to be significant. First, with training on fast speed of reaction systems for all *S*s involved in the comparison, those *S*s tested on the small moment of inertia system did better than those *S*s tested on the large moment of inertia system ($P < .01$). Second, with testing on large moment of inertia systems for all *S*s involved in the comparison, those *S*s previously trained on a slow speed of reaction system performed better than those *S*s previously trained on a fast speed of reaction system ($P < .05$).

DISCUSSION

The results showed clear differences in performance on a depth-keeping task favoring systems with fast speed of reaction over slow speed of reaction on both the training and test trials.[4] The

[4] As no data on amplitude of stick excursion was collected, it is possible that these differences can be attributed, in part, to differences in system responsiveness. Informal observations made by *E*, however, suggest

differential effect of moment of inertia was not demonstrated as a main effect. It was shown, however, that moment of inertia tested on did interact with the speed of system reaction trained on.

In terms of task difficulty, these results show that dynamic systems representative of slow reacting systems are more difficult to control than those more indicative of fast reacting systems. The data also indicate that training on fast reacting systems, the "easier" task, produces performance decrements when the system controlled during transfer is characterized by large moment of inertia effects. Taken together, these results show that the most difficult system to control possesses features which combine both large inertial effects and slow speed of system reaction; the LS system. It is interesting to note that the coefficients reflecting these two characteristics produce the longer control lags. Thus, these findings tend to support the generally accepted notion that systems with long control lags are more difficult to control than those with short control lags.

As for the question of which system to train on to produce the best transfer effects to all other systems, the data show that prior training on the LS system produced the greatest positive transfer effects regardless of which of the four systems the test was made on. That is, training on the LS system resulted in no appreciable decrement on either the SF, SS, or LF systems, and did produce significantly better performance on an LS system than did prior training on any of the other systems.

These findings tend to support the general conclusion drawn by Day (1956), that transfer from a difficult task to an easy task is better than the reverse order training.

that *S*s did not use full stick displacements after the first few training trials since in most cases, being naive, this behavior was unrewarding. Experienced operators can compensate for the slowness of a system reaction, but this is more likely to occur on a depth-seeking rather than a depth-keeping task.

Summary

Four groups of 20 *S*s were trained on a complex tracking task differing in the degree of control lag. Following training, 5 *S*s from within each group were transferred to the other systems, with one subgroup transferred to the same system. It was found that, on the basis of mean time on target scores per five-trial block, the more difficult tasks were those with the longer control lags. It was also found that training on the most difficult task, the system with the longest control lag, produced generally better transfer effects than did training on any other system.

References

Day, R. H. Relative task difficulty and transfer of training in skilled performance. *Psychol. Bull.*, 1956, **53**, 160–168.

Holland, J. G., & Henson, J. B. Transfer of training between quickened and unquickened tracking systems. *J. appl. Psychol.*, 1956, **40**, 362–366.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

Lordahl, D. A., & Archer, E. J. Transfer effects on a rotary pursuit task as a function of first task difficulty. *J. exp. Psychol.* 1958, **56**, 421–426.

Ritchie, M. L., & Michael, A. L. Transfer between instrument and contact flight training. *J. appl. Psychol.*, 1955, **39**, 145–149.

(Received January 11, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 376–380

EFFECTS OF EXPOSURE TIME AND DENSITY ON VISUAL SYMBOL IDENTIFICATION [1]

WARREN H. TEICHNER AND ERNEST SADLER

University of Massachusetts

In a previous study (Teichner, Reilly, & Sadler, 1961) *Ss* were presented for 1 sec. with visual displays containing randomly sampled, randomly positioned groups of alphabetic letters. One group reported the number of different letters displayed (counting group); another group reported the names of the letters (naming group). The number of letters named by the naming group was then counted and compared with the number reported by the counting group. It was found that the number counted was more accurate in estimating the number displayed and that as the number of letters displayed increased, the error of the naming group increased the more rapidly. Since letter detection and discrimination were common to both groups, but the memory requirement of the naming group was much greater, these results indicate: (*a*) visual symbol identification depends on both perception and short-term memory, and (*b*) perceptual capacity is greater than short-term memory capacity.

Although both perception and short-term memory contribute to the accuracy of symbol identification, their contributions may not always be in the same direction. On the one hand, for randomly positioned letters, as the total information on the display increases, the average amount of information per fixation during scanning will also increase and this should increase the probability of letter detection since it permits more letters to be detected in fewer fixations. Up to that exposure time which allows complete scanning of the display, the perceptual process should be aided as the number of letters increases. On the other hand, findings are available (Woodrow, 1938) which indicate that there is a mutual interference of letters when they are presented in groups and this interference increases as the number of letters increases and as the distance between letters decreases.

Woodworth and Schlosberg (1954) refer the Woodrow effect to differences in the fidelity of letter images falling at or near the fixation point and those falling in peripheral vision. Although this is not unreasonable as a visual phenomenon, by the same reasoning as before, as the number of letters on the display increases, the average number of letters near the fixation point will increase and again the probability of letter detection should increase. Thus, perception should be aided rather than handicapped. But, it has been shown (Teichner et al., 1961) that the percentage of correctly identified letters decreases under these circumstances and the decrease is at least as much associated with the omission of letters as it is with incorrectly identified letters. A better explanation might relate the mutual inhibition effect to the process of storage in short-term memory (Sperling, 1960; Teichner & Myers, 1961), which would tend to offset the expected perceptual gain.

As exposure time increases, the required rate of memory storage decreases and memory should be aided. Beyond some time permitting adequate scanning of the display, perception should not be aided. For a constant exposure longer than this scanning time, increases in the

[1] Presented to the Psychonomic Society, Chicago, September 1960. Performed under Contract N61339-588 with the United States Naval Training Device Center, Port Washington, New York.

number of symbols on the display should, therefore, produce a memory loss without any change in perception. The result should be an increasing loss in identification as the number of symbols (amount of information) on the display increased. Conversely, it should be possible to decrease exposure time to a point where the required storage rate is so high that recall will be a minimum. Under this condition, increases in display information should produce a negligible increased memory loss since there will be so little left to be lost. However, as noted above, the same increases will aid perception. The result of this should be an increase in identification as display information increases. Thus, it can be predicted that for very short exposure times, the accuracy of identification will increase, the greater the number of symbols to be identified, and that for longer exposure times, the reverse will be true.

Since the absolute number of correct identifications should increase with increasing display information, there being more symbols to be identified, a test of the predicted interaction between number of symbols and exposure time could be contaminated by a confounding of this type of effect with the interaction effect itself. It is desirable, therefore, to introduce the required perceptual and memory effects at constant information levels. This can be achieved by varying the density of the letters, i.e., by repeating the letters randomly at different parts of the display while keeping the number of different letter categories constant, and requiring that *S* report letter categories without regard to density. As letter density increases, the average proportion of the total information contained within a single fixation will increase; at the same time the number of letters will increase and their spacing will decrease, thus increasing the negative memory effect. Then, the following specific predictions can be made about the interaction of exposure time and density at any constant level of display information content: (*a*) At very short exposure times, identification will be positively related to density. (*b*) At relatively long exposure times, identification will be inversely related to density. (*c*) At some intermediate exposure time, density will have no effect. (*d*) At any density level, identification will be directly related to exposure time. The purpose of this experiment was to test these predictions at different levels of information content of the display.

Method

The methods used were similar to those of Teichner, Reilly, and Sadler's (1961) naming group. The display slides varied from 3 to 6 in number of letter categories and from 1 to 5 in letter density. There were 20 combinations of density and category, each of which was replicated 10 times with a different random sample of the alphabet to make up 200 slides. The time between slides was 10 sec. Two exposure times selected from 0.5, 1.0, 1.5, 2.0, 2.5, and 5.0 sec. were presented in successive halves of each session. The order of exposure times throughout the experiment was independently randomized for each of three groups of 4 *Ss* each.

A 30-min. preliminary session was devoted to instructions and practice with a sample of slides not used later and with different exposure times selected from the same time range. The experimental sessions were conducted on the 3 following days. In those sessions all 200 slides were presented in random order twice, once for each of the two exposure times of the session. A different random order of slides was used for each exposure time. Each session lasted about 1 hour depending on the combination of exposure times involved. A short rest was always given between each exposure condition.

The *Ss* were experimentally naive, volunteer, undergraduate students in competition for prizes of $20, $10, and $5 for the highest scores summed over all experimental sessions. No knowledge of results was given until the entire experiment was completed.

Results

The data were analyzed in three ways: (*a*) in terms of the percentage of displayed categories which were correctly identified, (*b*) in terms of the percentage number of categories in *S*'s report less than the number displayed (error of omission), and (*c*) in terms of the percentage number of categories in the report greater than the number displayed (error of commission).

Errors of commission were essentially zero at a density of 1 and for greater densities never reached a value as large as 10%. Since there was also no consistent effect on these errors of exposure time, they were rejected for further analysis. Errors of omission were inversely related to percentage correct. For this reason only percentage correct data are reported.

Table 1 reports that an analysis of variance of the percentage of correct identifications found the main effects of exposure time, categories, and density all significant at the .01 level.[2] The first and second order interactions of these variables were also significant at the .01 level except those involving *Ss*.

TABLE 1

Analysis of Variance of Percentage Correct Identifications

Source	*df*	*MS*	*F*
Between *Ss*	11		
Within *Ss*	1428		
Exposure Time (T)	5	4.510	50.11*
Categories (C)	3	3.880	485.00*
Density (D)	4	.050	6.25*
T × C	15	.050	8.33*
T × D	20	.028	4.67*
C × D	12	.028	3.50*
T × C × D	60	.005	1.67*
S × T	55	.090	
S × C	33	.008	
S × D	44	.008	
S × T × C	165	.006	
S × T × D	220	.006	
S × C × D	132	.008	
S × C × D × T	660	.003	
Total	1439		

* $P < .01$.

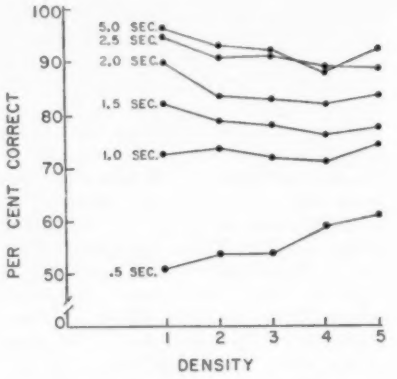

Fig. 1. Percentage correct as a function of density with exposure time as a parameter.

Figure 1, which presents the effects of density and exposure time pooled over categories, shows that accuracy increased with increasing exposure time up to at least 2.5 sec. The small difference between the 2.5-sec. and 5.0-sec. times suggests a limit somewhere between these two values. When these data are replotted on a time base, percentage correct is a negatively accelerated increasing function of exposure time with an asymptote of approximately 2.5 sec.

Figure 1 also reveals the nature of the interaction between exposure time and density. At the shortest exposure time percentage correct was essen-

[2] Complete data by *Ss* for all conditions have been deposited with the American Documentation Institute. Order Document No. 7035 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

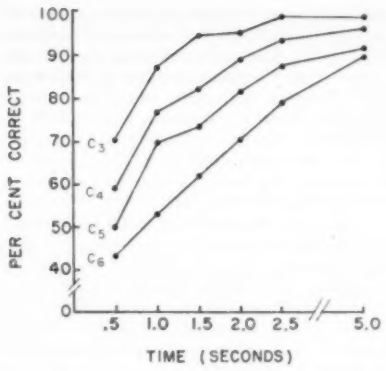

FIG. 2. Percentage correct as a function of exposure time with categories as a parameter.

tially directly proportional to density. At the 1.0-sec. exposure no general trend is apparent. At exposure times greater than 1.0 sec., there is a downward trend indicating an inverse relationship. The data suggest that the slopes of these downward trends tend to increase with increasing exposure time. The slopes are small in all cases.

Figure 2, which presents percentage correct as a function of exposure time, pooled over density, with number of categories as the parameter, shows that percentage correct was inversely related to number of categories and was a negatively accelerated increasing function of time. The asymptote of these time functions appears to decrease with increasing time; except for the six-category case, increases beyond 2.5 sec. are small.

Figure 3 presents percentage correct as a function of categories with exposure time as the parameter. This figure indicates that accuracy was in general inversely proportional to number of categories; the effect of exposure time was to reduce the slope of the function.

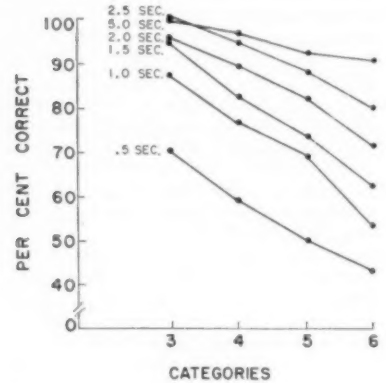

FIG. 3. Percentage correct as a function of number of categories with exposure time as a parameter.

DISCUSSION

The results appear to be as predicted. At the 0.5-sec. exposure time identification increased as density increased. This was predicted on the basis of a perceptual gain due to the increased proportion of the total information contained within a fixation and the consequent increase in the probability of letter detection. At this short exposure time, short-term memory was presumably approaching a lower limit; thus very little further loss would be possible as the effect of increasing interference. At the 1.0-sec. exposure time that effect was sufficiently important to nullify the perceptual gain and at longer exposure times, it overrode it. The data suggest that by 2.5 sec., exposure time was generally long enough to provide for adequate scanning so that further improvements in perception were small. On the other hand, the memory interference due to density was still present at the longest exposure time.

The results obtained are in good agreement with correlated points of the previous experiment (Teichner et al., 1961) which also found no density effect for the 1-sec. exposure time and which found the same kind of relationship between identification and the number of letter categories.

Density, as we are using the term, is a form of redundancy and in a noisy channel, redundancy would be expected to increase information transmission (Shannon & Weaver, 1949). On the other hand, in an essentially noise-free channel, as is the case here, there are psychological studies which show that redundancy in the form of density has a decrement-producing effect (Christner, Schutz, & Ray, 1959; Green & Anderson, 1956). Our study, then, is relevant to the explanation of these results and suggests that where redundancy increases information transmission there is a relatively large gain in the rate of transmission along a perceptual channel and otherwise there is a more important interference with short-term memory. This approach also suggests, for the same reasons, that the memory process is largely responsible for the loss in information transmitted which is commonly found with increases in stimulus information. Thus, the present study extends previous ones which have shown that *S* perceives more than he can recall (Sperling, 1960; Teichner & Myers, 1961; Teichner et al., 1961) by demonstrating the nature of the interaction between the two in their effects on identification.

Finally, regarding density, although redundancy is considered a means of increasing transmission along a noisy channel, so far as we know, it has not been suggested that redundancy might be a means of increasing transmission along an already overloaded but noise-free channel. Our results suggest that for the human channel and the special case of a short exposure time, this may be reasonable, at least for the form of redundancy used here.

Summary

Twelve *Ss* were presented with displays of alphabetic letters which varied in number of different letter categories, letter density, and exposure time. On the basis of a previously-made experimental distinction between perceptual capacity and short-term memory capacity, it was predicted that at short exposure times, increasing density would increase letter identification, but that at longer exposure times the opposite effect would be obtained. The results were in the direction of the prediction and in addition appeared to warrant the following specific conclusions for the situation studied:

1. Up to at least five symbols, percentage correct identification is a negatively accelerated increasing function of exposure time up to 2.5 sec. after which a limit is approached.

2. Percentage correct identification is approximately inversely proportional to the number of symbol categories (amount of displayed information).

3. The percentage of omitted symbols is a negatively accelerated decreasing function of exposure time with a limit of approximately 2.5 sec.

References

Christner, C. A., Schutz, R. G., & Ray, H. W. Some factors affecting visual search time for symbols on a large display. Paper read at American Psychological Association, Cincinnati, Ohio, 1959.

Green, B. F., & Anderson, L. K. Color coding in a visual search task. *J. exp. Psychol.*, 1956, **51**, 19–24.

Shannon, C. E., & Weaver, W. *The mathematical theory of communication.* Urbana: Univer. Illinois Press, 1949.

Sperling, G. The information available in brief visual presentations. *Psychol. Monogr.*, 1960, **74** (11, Whole No. 498).

Teichner, W. H., & Myers, J. L. Training aspects of decision making. *USN Train. Dev. Cen. tech. Rep.*, 1961, NAVTRADEVCEN 588-1.

Teichner, W. H., Reilly, R., & Sadler, E. Effects of density on identification and discrimination in visual symbol perception. *J. exp. Psychol.*, 1961, **61**, 494–500.

Woodrow, H. The effect of pattern upon simultaneous letter-span. *Amer. J. Psychol.*, 1938, **51**, 83–96.

Woodworth, R. S., & Schlosberg, H. *Experimental psychology.* New York: Holt, 1954.

(Received January 13, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 381-386

SENSORY INTEGRATION WITH AND WITHOUT REINFORCEMENT [1]

VERNON O. TYLER, JR. [2]

University of Nebraska

Sensory integration or S-S learning has been a notion of interest for some time, but it has not yet received adequate empirical support (MacCorquodale & Meehl, 1954, p. 213; Spence, 1950). Osgood (1957) cites evidence for the concept, but is not fully satisfied with these findings. The purpose of the present study was to test the notion once again, but with a more rigorous methodology.

Osgood postulates three kinds of learning: S-R, R-R, and S-S. In his theory, S-R learning is a function of a number of factors, with reinforcement an important one; S-S and R-R learning are said to be relatively independent of reinforcement with contiguity postulated as sufficient for the establishment of associations.

In this discussion, S will be used to represent within-the-organism stimulus correlates or traces resulting from the presentation of a stimulus and R to represent within-the-organism precursors of responses or traces of these precursors.

To demonstrate S-S learning, it is necessary to present stimuli contiguously in order to ascertain whether or not associations between their corresponding Ss are formed. At the same time, however, it is necessary to make certain that no S-R or R-R associations are formed or strengthened. This can be done by preventing responses from occurring or by making certain that overlearning has previously taken place so that any existing S-R associations cannot be strengthened.

Following the contiguous presentation of a number of stimuli, the presence of sensory integrations among their stimulus correlates may be ascertained by presenting some of the stimuli and observing whether or not there is evidence that the correlates of the remaining stimuli are activated in some way. At the present time, the only tangible evidence that these remaining stimulus correlates have in fact been activated is by the occurrence of some signifying or indexing responses linked to these correlates.

Sensory integration then will be defined here as the associations which develop among the internal stimulus correlates of a group of stimuli (or parts of a stimulus complex) which are presented together such that a part of the group, in a later presentation, will elicit a set of indexing responses which have occurred previously only to the total original group of stimuli. This definition is essentially the same as Osgood's, although it is more explicit.

This procedure poses several methodological problems: (*a*) Since the indexing responses must be suppressed during the training phase of this procedure, it is difficult to ascertain that the stimuli whose correlates are to be integrated do in fact impinge upon the receptors in an effective manner. For example, in the visual modality, *E* must be sure that *S* actually "sees" the stimuli. If, in a test for sensory integration, positive results are obtained, it would appear ob-

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Nebraska. The author thanks Katherine E. Baker, dissertation supervisor, for her help in the formulation of this design and in seeing this research to completion. The author also thanks John C. Loehlin for his valuable insights and stimulation on a number of aspects of this work.

[2] Now at Fort Worden Diagnostic and Treatment Center, Port Townsend, Washington.

vious that the stimuli must have impinged upon the receptor involved. Negative results, however, would leave this question unanswered. (*b*) Another problem is that of controlling the indexing responses during training. This is necessary in order to exclude S-R and R-R interpretations of positive findings. This would mean making certain that no new associations between Ss and Rs (S-R learning) or between Rs (R-R learning) involving indexing responses were developed during S-S training. Solution to this problem could involve using indexing responses each one of which has an overlearned association with one and only one stimulus and little or no association with any other stimuli, and by having *S*s make competing responses to the stimuli during the training. (*c*) Another problem is that of developing a sensitive means of measuring the indexing response during the test phase of the experimental procedure. In this experiment the tachistoscope was used because it appeared well suited for controlling the amount of "information" received by *S* over a considerable range and in sufficiently fine gradations to make a test sensitive even to minute variations in the strength of S-S associations.

It is the failure to control responding which leaves most of the previous work in this area inconclusive. In Solomon and Postman's (1952) study, responding was clearly present with *S* pronouncing nonsense words. The same is true of Cohn's (1954) work. In Gibson's studies (Gibson, Walk, Pick, & Tighe, 1956) in which animals had prolonged exposure to visual forms, responses while not differentially reinforced were certainly not rigidly controlled. Nor does Postman and Conger's (1954) study fully control the recurrence of a response—certainly not with laboratory precision.

The primary hypothesis tested in this study was that tachistoscopic thresholds are lower for recognition of nonsense words to which *S*s have been previously exposed (with indexing responses absent during training) than for nonsense words which are new to them (control words). It was also hypothesized that thresholds are independent of reinforcement during training, at which time S-S integrations are presumably being formed.

Procedure

Word lists and apparatus.—A large pool of five-symbol, nonsense "words" consisting of four capital letters and one Arabic numeral was constructed. The letters were randomly assigned to the words so that no letter was repeated until the alphabet was exhausted and no letter was repeated in any one word. The third symbol of each word was a randomly assigned digit from 1 to 9. These nonsense words were typed on a paper tape which was fastened to rollers in a Gerbrands mirror-type tachistoscope.

The tachistoscope was employed to check the visibility of the words. The ascending method of limits with .01-sec. steps was used to determine the duration of exposure necessary for *S*s to identify correctly four out of the five symbols (letters or digits) making up each nonsense word, ignoring any additional symbols which might be erroneously guessed to be present. This lower-than-usual criterion was used because it permitted greater sensitivity in detecting small differences in visibility of words. The illumination in the tachistoscope as measured by a Macbeth Illuminometer was 1.3 apparent ft-c at the eyepiece for both the "blank" field and the stimulus exposure field. The tachistoscope was situated in a small experimental room illuminated by a 100-w. red light bulb to minimize loss of dark adaptation when *S* looked away from the eyepiece.

From preliminary experimentation, a number of words were selected for which approximately equal thresholds were obtained. From this list, two pairs of words were selected so that the members of each pair were as nearly equal as possible in means and ranges of the thresholds they yielded. The four words of these two pairs were made into a final list ordered on the basis of obtained thresholds beginning with the word giving the lowest average threshold and ending with the word giving the highest average threshold. This order was chosen with the object of obtaining fairly uniform thresholds by balancing out the tendency for thresholds to decrease with ordinal position in the list (Howes & Solomon, 1951) and individual word differences in threshold based on their inherent differ-

TABLE 1

LIST USED FOR OBTAINING THRESHOLDS AND USE OF EACH WORD FOR TWO GROUPS OF *Ss*

Words	Group I	Group II
1. NA8XV	Practice	Practice
2. GT7BZ	Practice	Practice
3. LQ2UZ	Practice	Practice
4. IA9SJ	Practice	Practice
5. WY9OQ	Experimental	Control
6. GW6QB	Control	Experimental
7. YA5TI	Experimental	Control
8. LX7OJ	Control	Experimental

ences in visibility. In addition, four high-threshold practice words were placed at the head of this list, which is presented in Table 1.

This list was presented to all *Ss* in the main experiment. Words 5 and 6 comprised one pair of equated words and 7 and 8 another pair. Words 5 and 7 were the experimental words for Group I with Words 6 and 8 serving as controls. For Group II, Words 6 and 8 were experimental words and 5 and 7 served as control words. The use of this counterbalanced design precluded attributing any positive results to differences in word difficulty or serial position.

Preliminary training involved presenting two of these words together with other "filler" words. For this training, two sets of 60 3 × 5 in. training cards were prepared. A nonsense word was typed in the center of each card. The 60 cards of Set I (presented to *Ss* of Group I) consisted of 25 cards with Word 5 (from the list of Table 1) typed on them, 25 cards with Word 7, and 10 cards, each with a different word, which served as the fillers. Each filler card had 1 of 10 nonsense words randomly drawn from the original pool. The 60 cards of Set II (presented to Group II) consisted of 25 with Word 6, 25 with Word 8, and 10 with the same filler words as Set I. The cards of each set were thoroughly shuffled before every administration.

Subjects.—Twenty-eight *Ss* participated in the main experiment, 16 in Group N (neutral) (3 men and 13 women), 12 in Group R (reinforcement) (8 men and 4 women). Both of these groups were in turn divided into halves, Groups I and II, in order to counterbalance experimental and control words as described above. Most of the *Ss* were high school students who volunteered for the experiment while a few others were volunteers with whom *E* was acquainted. Age varied from 13 to 22 yr., with most of the *Ss* between 15 and 19 yr.

Procedure.—The general procedure was to present one of the sets of cards to an *S*, a card at a time, and then take him to another room and obtain his tachistoscopic thresholds. The *Ss* from both Groups N and R were randomly assigned to Group I or II and presented the cards of the corresponding set, I or II. The *S* was taken to a small, well-illuminated experimental room with a table and two chairs. A wooden screen on the table separated *S* from *E* so that he could not be observed "recording data." The instructions, paraphrased, were as follows:

> In this research, we are interested in how clearly things appear to people. This is important in such things as making instrument panels for aircraft and automobiles. Take this stylus (*E* hands stylus to *S*), point to each letter here in turn, and say whether it is "round" or "straight." That is, if a letter is made up of round lines or a combination of round and straight lines, call it "round." If it is made up of only straight lines, call it "straight." Just move the stylus along saying "round" or "straight." Relax and go at your own speed.

It should be mentioned here that the type of response required of *S* was intended to to prevent his subvocalizing the names of the letters. Two pilot studies indicated that this procedure was effective. In one of these studies, for each of the letters in a series of nonsense words, each *S* was deliberately instructed to make the indexing response in a covert manner and then to make overtly the presumably competing response. The *S* was first instructed to "think" the names of each of the letters in the nonsense words and then say whether each letter was "round" or "straight." After he had gone through this procedure, he was asked to say *only* whether the letters were round or straight, going through the word. With 16 *Ss* the mean time (7.82 sec.) required for thinking names of letters *and* making competing responses was significantly greater than that required for making competing responses alone (4.42 sec.), using a *t* test for correlated measures ($P < .005$). These data indicate that the competing responses effectively prevented even the covert exercise of indexing responses.

In the second pilot study, it was also shown that indexing responses were not covertly rehearsed while *Ss* were making competing responses. This was done by demonstrating

that it takes *S* no longer to make responses such as counting or describing letters than it does to make the same type of responses to completely new, unfamiliar stimuli for which no old, overlearned response labels are available. If it takes as much time to make a given type of competing response to both familiar and unfamiliar stimuli, then it is doubtful that an *S* has "smuggled in" any implicit indexing response to the familiar symbols during training. This procedure was utilized with 8 *S*s and a comparison was made of the time required to describe letters as either round or straight and to describe nonsense figures which were equivalent to the letters with respect to the number of curved and straight elements. The mean time required for describing the two kinds of stimuli did not differ significantly (5.06 and 5.09 sec., respectively, $P > .90$), using a t test for correlated measures. These results were taken as further evidence that *S*s do not subvocalize names of letters while describing them under these training conditions.

Returning to the experimental procedure, it should be said that *S*s in Group R were given an approving word or phrase immediately after they completed their performance of the task of naming round and straight symbols on each card. These phrases, randomly used, were such expressions as "Good," "Excellent," "Fine," and "Well done."

After presentation of the cards, *S* was immediately taken to another experimental room and seated in front of the tachistoscope. He was told (paraphrased here):

> Here we are studying the effects of lighting on the visibility of letters and numbers. Now look into this eyepiece. Do you see the two horizontal lines? I am going to flash a word between them for a brief interval and I want you to guess how it is spelled. It is not an ordinary word but a nonsense word—just a random arrangement of letters and numbers. The first time I flash the word, you probably won't be able to see much—nothing more than a flash of light. However, I will keep flashing the word for longer and longer intervals until it will be easier to recognize the letters and numbers. Start guessing as early as possible, even though you're not sure.

Before each exposure, *E* said "Ready?"

Thresholds were then obtained on all eight words in the list, using the ascending method of limits. The tachistoscope was first set to flash a word for a duration of .01 sec. The second exposure was for .02 sec., and so on in .01-sec. increments. For the four practice words, this procedure was continued until *S* "spelled" out all five symbols of the word perfectly or until an exposure of .80 sec. was reached. Only 2 *S*s failed to spell a word correctly before this limit of .80 sec. This use of the 5/5 symbols criterion for the first four words was to give *S*s more practice and to emphasize accuracy before they reached the critical words.

Thresholds for the remaining four words (Words 5 to 8) were obtained with the same procedure except that the criterion was 4/5 symbols correctly guessed in the proper order, but with extraneous intervening symbols permitted.

Results and Discussion

This procedure yielded four threshold measures for each *S*, two for control words of no prior exposures and two for experimental words with 25 prior exposures each. The scores for the two control words for each *S* were arithmetically averaged and, likewise, scores for the two experimental words for each *S*. Thus each *S* contributed two scores (two averages) to the statistical treatment. Means and *SD*s for each of the four experimental conditions are presented in Table 2.

It is apparent from these data that (*a*) both Groups N and R had lower thresholds for experimental words than for control words; (*b*) Group R had lower thresholds than Group N for both experimental and control

TABLE 2

Mean Thresholds (Sec.) for Four Experimental Conditions

Group	Control Words (Frequency = 0)		Experimental Words (Frequency = 25)	
	Mean	*SD*	Mean	*SD*
Neutral	.1413	.036	.1194	.041
Reinforcement	.1221	.036	.1054	.030

Note.—Original thresholds were in .01 sec., *S* means in .001 sec., and Group means in .0001 sec.

TABLE 3

ANALYSIS OF VARIANCE OF THRESHOLDS

Source	*df*	*MS*	*F*
Between *Ss*	27		
Reinforcement (R)	1	.00376202	1.739
Error (b)	26	.00216382	
Within *Ss*	28		
Frequency (F)	1	.00540175	8.093*
R × F	1	.00009304	<1
Error (w)	26	.0006675	

* $P < .01$.

words; and (*c*) the slopes of the curves for the two groups, if plotted graphically, would be essentially the same.

To evaluate the data statistically, they were subjected to the analysis of variance presented in Table 3. The predicted Frequency effect (experimental vs. control words) is highly significant. This indicates that exposure to the nonsense words (experimental words) reduced tachistoscopic thresholds for their recognition significantly below thresholds for words *Ss* had not seen before (control words). Thus the primary hypothesis of this study postulating a sensory integration type of learning is supported. In addition, the direction of the difference of the Frequency effect in these data supports the prediction that tachistoscopic thresholds are a decreasing function of the frequency of exposure; or, in the general terms of the theory, they support the hypothesis that the strength of sensory integration is an increasing function of the frequency of contiguous presentation of parts of the stimulus complex.

Although the Reinforcement effect is not statistically significant, there is a trend in the direction suggesting that reinforcement was serving to raise the general level of *motivation* of reinforced *Ss*, resulting in lower thresholds on all words (experimental *and* control).

The effect of reinforcement on learning, as opposed to performance, is indicated by the Frequency × Reinforcement interaction. It is clear there is no such interaction effect. This indicates that while Group R performed somewhat better (i.e., had lower thresholds) than Group N, the superiority of Group R to Group N was about the same for both experimental *and* control words. Since only the experimental words in Group R were reinforced, the control words not being given during training when the reinforcement occurred, and since both experimental and control words in this group show about the same margins of superiority over the corresponding experimental and control words of Group N, it was concluded that reinforcement did not strengthen sensory integration. The fact that no significant interaction effect occurred supports the second hypothesis of this study, namely Osgood's assumption that reinforcement has no effect on the *strength* of sensory integration and probably is not required even for the *formation* of sensory integration.

In general, the results of this study appear to provide ample evidence for the validity of the notion of sensory integration or S-S learning. As far as can be discovered from the literature, they represent the first data that unequivocally show that sensory integration can be demonstrated experimentally under appropriately and adequately controlled laboratory conditions.

These findings throw new light on the question of whether a response is necessary for learning. It seems reasonable to conclude that no response is necessary for at least one kind of learning. This conflicts with Hullian theory which utilizes only S-R learning and supports Tolman in his espousal of S-S learning.

Moreover, these findings raise once again a point that Tolman has stressed for some time. Perhaps many experimental situations which have been taken as instances of pure S-R learning may best be interpreted as S-S learning subtly contributing to responding. Now that greater attention has been given to controlling responding, typical S-R situations can be restudied to see what aspects of performance changes might take place when responding is controlled during preliminary training. This does not say that S-R conceptualizations are going to be eliminated by this approach, but rather it points out the need to delimit the area to which S-R notions alone are sufficiently explanatory. If different kinds of learning do follow different laws, certainly this calls for mapping their particular spheres of functioning in the organism.

Summary

The purpose of this study was to demonstrate S-S learning with particular care to control responding and thereby exclude S-R and R-R interpretations of the results. In the training phase of the experiment, *Ss* were asked to point to the letters of nonsense words and describe them in a way that made certain that the names of the letters were not uttered vocally or subvocally, thus ruling out S-R and R-R interpretations. In the test phase, tachistoscopic duration thresholds were obtained for the nonsense words and equated control words.

Thresholds were significantly lower for previously exposed words than for control words, which was interpreted as evidence for S-S learning. One group of *Ss* was praised during the training phase and did no better on the tachistoscope than an unpraised group. This was interpreted as indicating that S-S learning is independent of reinforcement.

References

Cohn, B. The effects of frequency and consequence as revealed by several different measures of verbal learning. *J. abnorm. soc. Psychol.*, 1954, **49,** 290–297.

Gibson, E. J., Walk, R. D., Pick, H. L., & Tighe, T. J. Effects of prolonged exposure to visual patterns. *J. comp. physiol. Psychol.*, 1956, **49,** 239–242.

Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 1951, **41,** 401–410.

MacCorquodale, K., & Meehl, P. E. Edward C. Tolman. In W. K. Estes, S. Koch, K. MacCorquodale, C. G. Mueller, Jr., W. N. Schoenfeld, & W. S. Verplanck (Eds.), *Modern learning theory.* New York: Appleton-Century-Crofts, 1954. Pp. 177–266.

Osgood, C. E. Motivational dynamics of language behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1957.* Lincoln: Univer. Nebraska Press, 1957. Pp. 348–424.

Postman, L., & Conger, B. Frequency of stimulation and frequency of response as factors in recognition. *Science,* 1954, **119,** 671–673.

Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43,** 195–201.

Spence, K. W. Cognitive versus S-R theories of learning. *Psychol. Rev.*, 1950, **57,** 159–172.

(Received January 13, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 387-390

A PSYCHOPHYSICAL STUDY OF HUNGER IN THE RAT [1]

ROBERT C. BOLLES

Hollins College

There is a growing tendency to attribute motivational phenomena to the *stimulus* consequences of deprivation (Estes, 1958). One implication of this viewpoint is that it ought to be possible to do psychophysical experiments with the alleged deprivation-produced stimuli. It is a source of some little embarrassment therefore that drive discrimination and especially drive-intensity discrimination have been typically so difficult to demonstrate (Bloomberg & Webb, 1949; Jenkins & Hanratty, 1949). The purpose of the present study was to demonstrate that the difficulties encountered in previous studies are mainly methodological, and that an experimental psychophysics of hunger is feasible. Specifically, animals were trained to discriminate between large and small prefeedings; the discrimination was made increasingly difficult so that "just noticeably different" prefeeding values could be determined; and finally, the obtained jnd's were added to derive a psychophysical scale of hunger.

METHOD

Subjects.—The *Ss* were 54 male albino rats, approximately 100 days old at the start of training. Fifty *Ss* survived the 3 mo. of training and testing.

Apparatus.—Each *S* was confined for the course of its training and testing in an individual unit which consisted of a home cage, two alleys leading off the home cage in opposite directions, and goal boxes at the ends of the alleys. The home cage was 10 × 10 in. and 4 in. high, with hardware-cloth over sawdust flooring, and sliding doors closing off the alleys. The alleys were 4 in. wide, 4 in. high, and 4 ft. long and they terminated in goal boxes that were 4 in. wide, 4 in. high, and 10 in. long. Nonretrace doors separated the goal boxes from the alleys. The whole unit was made of unpainted wood, except for one side which was covered with hardware-cloth. Thus, *S* was provided with good spatial-orientation cues, since the two goals were well-separated in space, and since the two alleys presented extramaze cues on opposite sides of *S* as it ran. Leeper (1935) and Bolles and Petrinovich (1954) have indicated that good spatial-orientation cues can greatly facilitate drive discrimination. Twelve of these units were stacked together so that as many as 12 *Ss* could be tested at a time.

Training.—Throughout the study hunger and also, presumably, the to-be-discriminated stimulus consequences of hunger were controlled by prefeeding *S* a certain proportion of its average daily consumption. This procedure has the advantage over the more usual procedure involving deprivation for a certain number of hours since it provides for more rapid onset of the relevant stimuli. Prior to the discrimination training *Ss* had had a few days of 23-hr. deprivation experience, and had been given 2 days in which to explore the apparatus, 1 day with access to ad lib. food on one side, and 1 day with access to ad lib. food on the other. During training *S* was prefed in its individual home cage, where it had been confined for approximately 23 hr. When the prefeeding ration had been eaten (within 20 min., ordinarily) the food dish was removed from the home cage and *S* was detained there another 5 min. Then the sliding doors leading to the two alleys were withdrawn and *S* ran its one trial for the day. Food to finish the daily meal was in one goal box on days when the prefeeding ration was large, resulting in "low hunger," and in the other goal box on days when the prefeeding ration was small, resulting in "high hunger."

Large and small prefeedings differed between *Ss* to some extent because prefeedings

[1] The present study was made possible by the continuing support of the National Science Foundation. Preliminary findings were reported in an unpublished bachelor's thesis, by Henry Calam, Princeton University, and at the 1959 meeting of the Midwestern Psychological Association.

were only measured to the nearest .5 gm., and there was some variability among *Ss* in total consumption, but the large prefeeding amounted to about seven-eighths, and the small prefeeding to about one-eighth, of *S*'s average total daily intake. Thus, *S* was required to discriminate between high and low hunger where the difference represented about three-fourths of its daily food intake. High and low conditions and left and right sides were counterbalanced across *Ss*. Except as noted below, *Ss* were run under high and low hunger on alternate days. The "correction method" was used, so *S* could retrace after an error. An error was counted only if *S* ran far enough down the wrong alley to be able to see into the empty goal box, i.e., as far as a low barrier placed 12 in. from the end of the goal box. After entering the correct goal box, *S* was confined there until it had finished eating. The time of confinement stabilized during the course of training as *S*'s goal box consumption stabilized at some value appropriate for the size of its prefeeding. The goal box food dish always contained an excess of available food, and food was weighed regularly before and after eating so *S*'s gradually increasing consumption could be taken into account in preparing constant proportion prefeedings. A wet mash consisting of two parts water to one part ground Purina lab chow was used. A water bottle was always available in the home cage but *Ss* drank very little, rarely as much as 5 cc per day, because of the high water content of the mash. After the one daily trial, the doors were opened and *S* was allowed to run back to its home cage where it was confined without food until the next day. Discrimination training was continued until *S* met a 9 out of 10 criterion of learning.

Testing.—When the discrimination was established *S* was run on successive blocks of 10 trials where the difference between large and small prefeedings was progressively reduced. Thus, the discrimination was made progressively more difficult until, finally, as the prefeedings were made the same size, the task became impossible. Ideally, the discrimination performance would reflect the difficulty of the task, dropping from an initial level of 90% correct to a chance level, thereby yielding a psychometric function. The prefeeding differential at which performance was 75% could then be defined as a prefeeding jnd, a discriminability threshold.

Six different groups of *Ss* were run to determine discriminability thresholds at six different points on the prefeeding continuum. Thus, for one group large and small prefeedings were brought progressively toward one-half the daily consumption in order to determine the threshold at the middle of the prefeeding scale. Two groups of *Ss* were run to obtain threshold determinations at the extreme ends of the scale, and three groups were run to fill in intermediate points. Except for the variation in size of prefeeding, the same procedures were followed that had been used during the initial training.

Several precautions were taken to insure that *S*'s correct responding was contingent upon discrimination of its hunger, or at least upon the size of its prefeeding, rather than upon some cue inadvertently provided by *E* or the test situation. For example, the daily alternation of correct sides was frequently interrupted. Although *Ss* in a group were run within a few minutes of each other, the order of running them was constantly changed. As a consequence *E* typically did not know until *S* had made its daily run which side was correct.

Results

The median *S* (of the total 54) met the 9 out of 10 criterion of learning in 43 trials, and 33 *Ss* learned the discrimination within 60 trials. These 33 *Ss* then served under the six different sets of conditions used to determine discriminability thresholds at six different points on the prefeeding continuum. Either 5 or 6 *Ss* were tested under each set of conditions. Figure 1 gives the results for one group of *Ss* for which large and small prefeedings were brought progressively closer to 50% of the mean daily consumption. Each point on Fig. 1 represents the performance on a block of 10 trials of an individual *S* under a particular prefeeding differential. The figure is drawn as though all *Ss* were tested under the same conditions of differential prefeeding. Because prefeeding rations were only measured to the nearest .5 gm., and because there was some individual variation in total consumption, it was not possible to keep prefeeding differentials exactly the same for all *Ss*; but they were kept within about 5%,

and the median values are plotted in Fig. 1. The median percentage correct for the group at a given median prefeeding differential is also shown. The ordinate is plotted in normal deviate units so that the psychometric function could be drawn linearly to fit the group medians. The obtained discriminability threshold, the difference between "just noticeably different" values, amounted to 28% of the total daily food consumption.

For this group of *S*s high hunger had been reduced and low hunger had been increased symmetrically, so that, at threshold, *S* was just able to discriminate between a large prefeeding, amounting to 64% of its total daily food consumption, and a small prefeeding, amounting to 36% of its total daily food consumption. When the large and small prefeeding rations were equated, and made equal to 50% of the daily total, then the median performance for the group fell to 45% correct.

Discriminability thresholds were obtained at five other locations on the prefeeding (hunger) continuum. They varied in size from 40% (that is, the difference between 90% and 50%) toward the low hunger end of the continuum to 20% (the difference between 20% and 0%) at the high hunger end. These findings are summarized in Fig. 2. Taking the discriminability threshold, or jnd, as a unit of hunger intensity, a psychophysical scale of hunger may be derived in the classical Fechnerian manner: by adding up the empirically obtained jnd's and plotting them against their corresponding physical magnitudes (prefeeding rations). Figure 2 presents a smooth curve fitted to the present threshold determinations.

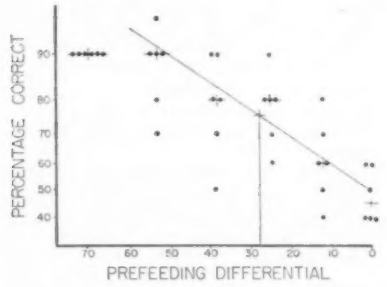

FIG. 1. Individual and group discrimination performance as a function of the size of differential prefeeding.

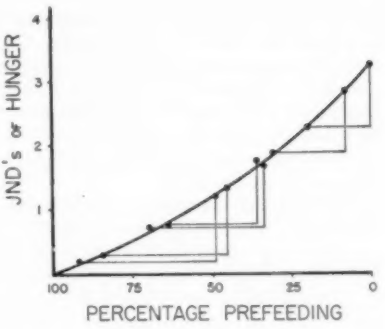

FIG. 2. Discriminability of hunger as a function of the percentage of daily food given in prefeeding.

DISCUSSION

Drive-intensity discrimination was obtained with far fewer trials in the present study than in the previous studies by Jenkins and Hanratty (1949) and Bloomberg and Webb (1949). This faster learning may probably be attributed to certain procedural advantages: (*a*) There were good extramaze cues for spatial orientation; (*b*) the onset and termination of the stimuli to be discriminated occurred at the choice point and in the goal box (Bolles & Petrinovich, 1954); (*c*) the onset and termination of the stimuli to be discriminated occurred quickly (achieved here by using a prefeeding procedure and feeding a wet mash that could be quickly consumed); (*d*) *S* lived at the choice point. (Preliminary investigation had shown that

if *S* was put into the choice point compartment just prior to prefeeding, then learning was retarded.) Further methodological refinements leading to smaller thresholds and simpler experiments may be hoped for, but the present procedures at least make drive-intensity discrimination technically feasible.

None of the control procedures gave any indication that *S*s were responding on any basis other than the size of the prefeeding ration. The best evidence that performance was contingent upon differential prefeeding was that in every case performance broke down on the impossible task. A few *S*s were put back on an easy discrimination task at this point. Some retained the discrimination, others had to be retrained. In short, there seems little doubt but that *S*s were responding discriminatively to the differential prefeeding.

Some doubt might arise about whether *S*s were actually hungry rather than thirsty, since they were receiving most of their daily water at the same time they were consuming their daily food. Upon the completion of testing, some *S*s were tested with water in one goal box and dry Purina pellets in the other. It was invariably found that water was refused but the dry food was accepted. Hence, it does not seem reasonable to suppose that *S*s were primarily thirsty. It should be recognized however that the particular results presented here in Fig. 2 might have been somewhat different if the food used had not been so wet.

The psychophysical function of hunger given in Fig. 2 is concave upward, in contrast with previous attempts to quantify hunger where curves convex upwards have generally been reported. The discrepancy may depend upon the fact that in the present study percentage prefeeding rather than the more usual hours of deprivation was the independent variable. Only further experimentation can determine whether these two experimental procedures consistently produce similarly discrepant quantitative effects. Similarly, further research will be necessary to establish whether the construct scaled and labeled "hunger" in Fig. 2 will have empirical correlates such that we will want to call it hunger. Another line of attack upon the problem of defining hunger might well start with a search for the discriminative stimuli in the present situation. Are they in the stomach, or the head, or somewhere else? Either approach to the problem might permit the much abused and ill-defined word "hunger" to be pinned down quantitatively in terms of the stimulus consequences of deprivation.

Summary

Rats were trained to respond differentially on the basis of small and large prefeedings. When the discrimination had become established (the median trials to criterion was 43) the task was made progressively more difficult by reducing the difference between small and large prefeedings. The point at which performance fell to 75% correct was defined as a discriminability threshold, and such thresholds obtained at different points on the prefeeding continuum were used to construct a psychophysical "scale" of hunger.

References

Bloomberg, R., & Webb, W. B. Various degrees within a single drive as cues for spatial response learning in the white rat. *J. exp. Psychol.*, 1949, **37**, 628–636.

Bolles, R., & Petrinovich, L. A technique for obtaining rapid drive discrimination in the rat. *J. comp. physiol. Psychol.*, 1954, **47**, 378–380.

Estes, W. K. Stimulus-response theory of drive. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1958.* Lincoln: Univer. Nebraska Press, 1958.

Jenkins, J. J., & Hanratty, J. A. Drive intensity discrimination in the white rat. *J. comp. physiol. Psychol.*, 1949, **42**, 228–232.

Leeper, R. The role of motivation in learning: A study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.

(Received January 14, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 391–395

KEEPING TRACK OF VARIABLES THAT HAVE FEW OR MANY STATES

DOUWE B. YNTEMA AND GAYLE E. MUESER

Lincoln Laboratory,[1] *Massachusetts Institute of Technology*

This article describes further experiments on keeping track of several variables. The *S* is read a series of messages, each of which tells him that one of the variables is, and until further notice will remain, in some particular state. At random intervals the series of messages is interrupted and *S* is asked what the present state of one of the variables is.

The problem considered here is how the fraction of the questions answered correctly depends on the number of states that each variable may assume. Keeping track of variables that have the same set of possible states has been shown to be different from keeping track of variables that have their own exclusive states (Yntema & Mueser, 1960); so it was decided that the effect of number of states should be examined in both situations. It was also decided that the probability that a message would change the variable to which it referred should be held constant while the number of possible states was varied. Preliminary experiments had shown that the task is especially difficult when a large fraction of the messages are changes.

METHOD

Three experiments were done to investigate the effect of number of states in a variety of cases. The experiments were much alike in plan, differing chiefly in the number of variables whose states *S* was to remember, and in whether the variables had the same or different possible states.

[1] Operated with support from the United States Army, Navy, and Air Force.

Experiment I

In Exp. I *S* attempted to keep track of six variables that had the same set of possible states.

Conditions and design.—Looking ahead at Fig. 1 will be helpful in remembering the conditions considered in the various experiments. As the top panel of Fig. 1 implies, Exp. I included 10 conditions: 2, 3, 5, 8, and 21 possible states, every message being a change; and 2, 3, 5, 8, and 21 states, the probability of change being one-fourth. Eight *S*s each made one run under every condition. The design governing the order in which conditions were presented to *S*s was balanced and was an incomplete latin square. Lists of stimuli for different *S*s were generated independently, and so were lists with different probabilities of change; but the lists presented to a given *S* in the conditions involving a given probability of change resembled each other in a way that will be explained below.

Lists of stimuli.—The variables were identified by the digits 1 through 6. In any one run they all had the same set of possible states, namely, n randomly chosen consonants. The list of stimuli for a run had three parts: the prestarters, the starters, and the body. The prestarters consisted of one message about each variable and were chosen by putting the variables in a random order and then choosing one of the n states for each at random. The starters also consisted of one message about each variable. The variables were again put in a random order and a state was chosen for each, but the probability of choosing the state that had been assigned to the same variable in the prestarters was now $1 - c$, where c was the probability of change, and each of the $n - 1$ other states had probability $c/(n - 1)$. The body of the list consisted of 128 items chosen one after another according to the following scheme: First, it was decided whether the item would be a message or a question. The probability of a question was one-half. Next, the variable to which the item would refer was chosen, all variables being equally likely. Finally, if the item was to be a message a state was selected. The probability that the state

would be the one named in the previous message about the chosen variable was $1 - c$, and the probability of each other state was $c/(n - 1)$.

There is some evidence that randomly generated lists of this sort differ in difficulty (Yntema & Mueser, 1960, p. 21). In order that such differences should not obscure the effect of number of states, the lists presented to a given S in the conditions involving a given probability of change were, unknown to S, made as nearly alike as possible. To be exact, the variables in any two such lists stood in one-to-one correspondence to each other in the following sense: if the ith item on one list was a question, then the ith item on the other list was a question about the corresponding variable; if the ith item was a message that changed a variable, then the ith item on the other list was a message that changed the corresponding variable; and if the item was a message that left the variable unchanged, then the corresponding item was a message about the corresponding variable leaving it unchanged. To reduce the possibility that S would learn the common structure of the lists, no variable was ever allowed to correspond to itself.

The lists were generated on an IBM Type 709 computer.

Apparatus and procedure.—The S sat before an inclined board to which were fastened six boxes, which represented the six variables whose states he was to remember. The boxes were labeled 1 through 6, and each contained a stack of blank 3 × 5 in. cards.

Before a run S wrote down the n possible states (unless all 21 consonants were possible) and posted them on a corner of the board where he could refer to them if he wished. He was told that messages would "almost always" change the boxes (if the probabilty of change was one) or would "usually not" change the boxes (if the probability was one-fourth). The E then read the prestarters. As each was read S wrote the consonant it contained on the top card in the appropriate box and put the card on the bottom of the stack out of sight. He was told he need not try to remember the prestarters.

After those preparations the task began. The E read the remaining items in the list one by one. Messages were spoken in the form, "Four B," and questions in the form, "Letter of Two?" When E read a message S repeated it aloud, wrote the consonant it contained on the top card in the appropriate box, and put the card on the bottom of the stack. As soon as the card was out of sight E read the next item. When E read a question S attempted to reply with the last consonant he had written in the box about which the question inquired, guessing if necessary. He was not told whether his response was correct. As soon as he replied E went on to the next item.

The S was instructed to work rapidly and not to "take the time to worry about the other boxes" while writing a message or answering a question. There were two runs to a session with a rest of about 25 min. between sessions. On Day 1 there were a practice session and two experimental sessions; the remaining experimental sessions were on Day 2. Before the practice session E and an assistant gave a short demonstration of the task. The practice lists were those used in Exp. II and will be described below.

The present method differs from that used earlier (Yntema & Mueser, 1960) in that S was prevented from seeing the old state of the variable when writing a message. The present arrangement seemed a little neater in an experiment directly concerned with the probability that the message would be a change; otherwise the difference in method is not considered important.

Subjects.—The Ss of Exp. I and II were 8 male Air Force lieutenants and 8 Air Force enlisted men. Four of each were chosen at random for Exp. I and the others were assigned to Exp. II. All 16 were operators of an experimental air-defense center. One lieutenant in Exp. I had been an S several months earlier in a similar experiment (Yntema & Mueser, 1960). All others were naive to experiments of this type.

Experiment II

In Exp. II there were again six variables (i.e., six boxes), but each now had its own set of n possible states. The description of Exp. I applies except where otherwise stated.

There were eight conditions, as may be seen in the middle panel of Fig. 1. The design for the order in which conditions were presented to Ss was a complete latin square.

The states of the first variable were called Directions. There were at most eight of them—north, northeast, east, etc. They were represented by arrows when S wrote them on the cards. The states of the other five variables were: Jewels—diamond, emerald, jade, opal, pearl, ruby, topaz, zircon—represented by their initial letters, written as capitals. Marks—check, plus, hat, stroke, comma, equals, star, question—"hat" represented by a caret, "stroke" by a line slanting down toward the right, "star" by three crossed lines, and the others by the usual symbols. Animals—bird, fish, cow, snake, giraffe, snail, swan, butterfly—represented by crude drawings

of the type used earlier (Yntema & Mueser, 1960). Countries—Austria, China, France, Germany, India, Mexico, Norway, Spain—represented by their initial capital letters. Numbers—1 through 8, written as digits. For a run in which every variable was to have n possible states, a subset of size n was drawn at random from each of these six sets. Before the run S wrote the symbols in the six subsets and posted them where he could refer to them if he wished.

The six boxes were labeled "Direction," "Jewel," "Mark," "Animal," "Country," and "Number." They were also labeled with a letter: K, H, L, Q, W, or X. The letter was changed from run to run, but was the same for all boxes in any one run. Its only purpose was to allow the messages and questions to be spoken in a form comparable to that used in practice and in the other experiments—i.e., in the form, "K north," and "Jewel of K?"

All *Ss* of Exp. I and II were given the same two lists for practice. The first was a half-length list with four variables: Animal and Country of K and X. The second was full-length and there were nine variables: Jewel, Mark, and Number of H, L, and W. In both lists every variable had four states and the probability of change was three-fourths. The *Ss* of the present experiment spent about 5 min. at the start of the practice session writing the 48 symbols and learning to associate them with the names of the states.

Experiment III

Except that there were three variables (i.e., three boxes) instead of six, Exp. III was essentially a repetition of the first two, and the description of Exp. I and II applies here except when otherwise stated. Some of the conditions examined in Exp. I and II were omitted, however. It was felt that if the probability of change were only one-fourth, the task of keeping track of three variables with different sets of states would be so easy that there would not be enough errors to analyze. Then too, keeping track of three variables that have the same two possible states is peculiar: S need only remember which variable if any is the odd one and what its state is. Those conditions were also omitted, since this was not intended as an experiment on variables whose states fall into patterns.

Twelve conditions remained, as may be seen in the bottom panel of Fig. 1. The states were Directions, Marks, and Animals when each variable had its own set. Lists for conditions that differed only in number of states resembled each other as in Exp. I and II, but with only three variables it was no longer possible to require that no variable should ever correspond to itself. The design for the order of conditions was a balanced incomplete latin square.

There were 8 *Ss*, all of whom had served in Exp. I or II about a month earlier. There were a lieutenant and an enlisted man from Exp. I, and four lieutenants and two enlisted men from Exp. II. They were chosen solely because they were still available. There was only one practice run, a full-length run with nine variables: Direction, Mark, and Animal of H, K, and L.

Results

Primary results.—The fraction of the questions answered correctly in each run was corrected for chance in the usual manner—that is, by solving for p the equation:

$$r = p + (1 - p)/n \qquad [1]$$

r is the raw fraction right.
n is the number of states each variable may assume.
p is the fraction right after correction for chance.

The correction was used because it made the results easier to describe—not because of any assumption that S guesses when he does not know the answer. It will be shown below that errors are not always random guesses.

Figure 1 shows the mean p for each condition. It may be seen that in none of the seven cases examined in these three experiments does the number of states have any large effect on the corrected fraction right. In fact, as far as these data can show the corrected fraction right is constant. Analyses of variance and nonparametric tests based on the coefficient of concordance (Kendall, 1955, Ch. 6) failed to show that the effect of number of states was significant at the .10 level in any of the seven sets of conditions.

Inspection of Fig. 1 makes it obvious that the probability that a message will change the variable to which it refers has a large effect. As

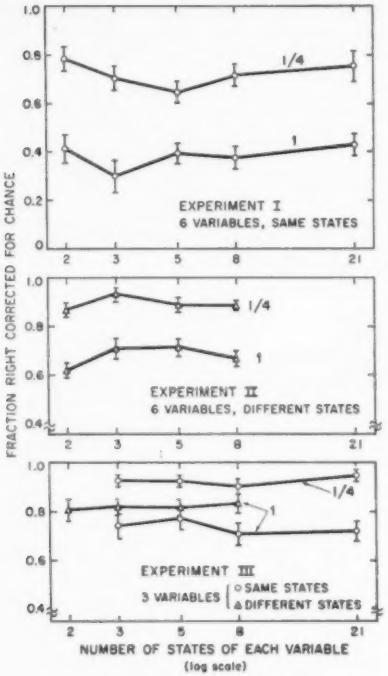

FIG. 1. Effect of n, number of states each variable may assume, on fraction of the responses correct after allowance for chance success. (The parameter is the probability that a message will be a change. Points are means for 8 *S*s. Vertical bars extend $\pm\sigma_m$.)

expected, *S*s keep better track of slowly changing situations than of rapidly changing ones.

Secondary results.—The corrected fraction right increased markedly from run to run in Exp. I, the average increase being about .02 per run with either probability of change. It was feared that the effect of number of states might be hidden by this order effect; so the linear component of the order effect was removed from the corrected fractions right observed in Exp. I and the tests for effect of number of states were repeated. The results were the same. The order effect helps to account for the unusually large standard errors shown in the top panel of Fig. 1, but its influence on the shape of the two functions shown there may be dismissed as negligible. Since the order of conditions was balanced, the linear component had no influence. There was no obvious nonlinear component; and since the design was eight-tenths of a latin square, only a quite obvious nonlinear component would have changed the ordinate of any point by as much as .01 relative to the grand mean.

An analysis was made of the frequency with which *S* responded with the previous state of the variable in question instead of its present state. Questions that were answered correctly were ignored. So were questions whose answers had been defined by a message in the starters or by a message that was not a change. The fraction of the remaining questions answered with the previous state was computed for each run with n greater than 2 and was corrected for chance by the formula obtained when $n - 1$

TABLE 1

CORRECTED PROBABILITY OF A RESPONSE WITH THE PREVIOUS STATE, GIVEN AN ERROR AND GIVEN THE LAST MESSAGE WAS A CHANGE

Type of Task and Probability of Change	Mean	σ_m
Six Variables:		
Same states, 1/4	.48*	.14
Same states, 1	.04	.04
Different states, 1/4	.52*	.14
Different states, 1	.57**	.08
Three Variables:		
Same states, 1/4[a]	.48*	.14
Same states, 1	.10	.06
Different states, 1	.58**	.08

Note.—Means are for 8 *S*s (except as noted), chance correction applied.

[a] 6 *S*s; the other 2 made no errors when the last message was a change.

* Different from zero at .02 level, two-tailed.

** Different from zero at .001 level, two-tailed.

is substituted for n in Equation 1. The corrected fractions did not appear to depend on n; so they were averaged over the runs made by a given S under conditions differing only in n. The means of the resulting averages are shown in Table 1. It may be seen that in five of the seven cases there was a significant tendency for S to respond with the previous state when he made an error. Why the other two cases should be exceptions is not entirely understood.

Discussion

These experiments agree with previous work on short-term recall of random, independent items (Gregg & Laughery, 1960; Lloyd, Reid, & Feallock, 1960; Miller, 1956; Pollack, 1953; Pollack, Johnson, & Knaff, 1959). Like the earlier results, the present ones can be interpreted as showing that the size of the set of alternatives from which the items are drawn has a relatively small effect on the average number of items that can be retained. In all of the previous experiments the data were collected or analyzed in such a way that chance successes probably did not have much effect on the results.

The improvement from run to run in Exp. I seems to be unusual: no important effects of order were observed in Exp. II or in an earlier experiment on naive *Ss* (Yntema & Mueser, 1960). It may have been a mistake to use digits for the names of the variables in Exp. I and consonants for their states. Combinations of digits and consonants are often meaningful: "2B" sounds like "to be," "4F" is a Selective Service classification, and many other pairs have special associations for an Air Force man. With practice the *Ss* may have learned to keep track of mediating responses like "to be" instead of keeping track of the consonants. If so, it is understandable that performance improved. The task is easier when the variables have different states (Yntema & Mueser, 1960); and a mediating response, being peculiar to one of the digits, is like a state that can be assumed by only one of the variables.

Summary

The task was to remember the present states of several variables. The *S* was read a series of messages, each telling him what the state of one of the variables was. He recorded the messages in such a way that he could not see them once he had written them. At random intervals the series of messages was interrupted and he was asked to recall what the last message about one of the variables had been.

The effect of the number of states that each variable might assume was examined in seven cases that differed from each other in one or more of the following ways: (*a*) there were three variables for *S* to keep track of at once, or there were six; (*b*) all variables had the same set of possible states, or each had its own distinct set of states; and (*c*) the probability that a message would change the state of a variable was one, or it was one-fourth. In none of the seven cases did the number of alternative states have any significant effect on the fraction of the questions answered correctly, provided the fraction correct was adjusted for chance success. The probability of change was found to have a large effect.

References

Gregg, L. W., & Laughery, K. R. Memory span for digits as a function of size of symbol set and mode of presentation. *Amer. Psychologist*, 1960, **15**, 478. (Abstract)

Kendall, M. G. *Rank correlation methods.* (2nd ed.) New York: Hafner, 1955.

Lloyd, K. E., Reid, L. S., & Feallock, J. B. Short-term retention as a function of the average number of items presented. *J. exp. Psychol.*, 1960, **60**, 201–207.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–97.

Pollack, I. Assimilation of sequentially encoded information. *Amer. J. Psychol.*, 1953, **66**, 421–435.

Pollack, I., Johnson, L. B., & Knaff, P. R. Running memory span. *J. exp. Psychol.*, 1959, **57**, 137–146.

Yntema, D. B., & Mueser, G. E. Remembering the present states of a number of variables. *J. exp. Psychol.*, 1960, **60**, 18–22.

(Received January 23, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 396–403

RECENCY AND FREQUENCY IN PAIRED-ASSOCIATE LEARNING[1]

LLOYD R. PETERSON, DOROTHY SALTZMAN, KENNETH HILLNER, AND VERA LAND

Indiana University

The assumption that repetitions build up associative strength by the accumulation of successive increments has recently been challenged (Estes, 1960; Estes, Hopkins, & Crothers, 1960). A single presentation of a list in experiments by these investigators appears to have produced little or no strength for some pairs of associates, while other pairs acquired high strength. The data have been interpreted as evidence for some form of all-or-none theory of the formation of associations.

An investigation of the time dimension may clarify some aspects of the formation of associations. The effect of a presentation must be assessed by some later measurement, and the interval elapsing between presentation and test may well be related to the number of correctly recalled associations. It seems likely that when perceptual conditions are favorable the probability of a correct recall might be close to 1.0 immediately following presentation. If a gradual decrease in the retention curve also occurs, a reasonable interpretation would be that associations are always formed but differ as to degree of retention.

An attempt to relate short-term retention to learning has been made by measuring recall of individual nonsense items after a single presentation over intervals up to 18 sec. (Peterson & Peterson, 1959). Recall was found to decrease very rapidly, with repetition significantly improving retention. This retention of items over short intervals would seem to relate directly to what Underwood and Schulz (1960) have called the "response recall" stage of list learning, during which time the items comprising the list are learned. Concurrent with this response learning in the list experiment, associations between stimuli and responses are established. The latter kind of cue learning has been investigated in a series of short-term retention experiments to be reported in this paper. Response learning was minimized by specifying in the instructions a set of numbers which were then used as responses repeatedly throughout a given experiment. The initial experiment investigated the retention function for associations over intervals less than the time between trials in the traditional list-learning experiment.

EXPERIMENT I

Method

Subjects and apparatus.—Twenty-four students from the introductory psychology course at Indiana University served as *Ss*. Participation in experiments was a course requirement. A memory drum was set in a 3 × 4 ft. black screen. A sliding shutter permitted exposure of any one of several columns of

[1] This research has been facilitated by National Science Foundation Grant G-12917, for which the senior author is principal investigator. The first author is responsible for the design, interpretation, and report of the investigation. The second author assisted in Exp. III and IV. The third author assisted in Exp. II. The fourth author assisted in Exp. I.

items. Twenty-eight exposures comprised a complete revolution of the drum.

Procedure.—A set of 108 three- and four-letter single-syllable words occurring at least 10 times per million in Thorndike and Lorge's (1944) count provided the stimuli. The responses were the digits 1–9.

The experimental session was divided into 12 blocks of presentations, each block containing 26 exposures of items, with a 4-sec. rest between blocks, save that between Blocks 6 and 7 a rest interval of 1 min. permitted change of the drum tape. A sample block is shown in Table 1. Nine words were selected at random from the previously described set to be used in a given block of exposures. A digit was randomly paired with each word. Five of these pairs of items were used to test the five experimental conditions, while the other four pairs were used as dummies to fill space. The first exposure in each block was a dummy to eliminate the advantage of initial position. Dummy pairs were presented and repeated where necessary to fill out blocks. The two members of a pair appeared together during a 2-sec. exposure, and recall was tested by presenting the stimulus alone after zero, one, three, or eight other exposures had occurred. Thus retention intervals of 0, 2, 6, and 16 sec. were tested. A repetition condition was also tested by presenting a pair twice in succession and then presenting the stimulus alone 6 sec. after the second presentation of the pair. For all conditions presentation and testing of one condition overlapped the retention cycle of other conditions. Pairs were randomly assigned to conditions within each block. A given word was used in only one block of the experiment, while the same set of digits was used again with new words in each succeeding block. The position of a condition within the block was varied randomly from one block to another. The *S* was instructed to pronounce aloud everything that appeared in the aperture. When only a word appeared, *S* was also to try to speak aloud the number last seen with the word. The *S* was instructed to guess when necessary.

TABLE 1

SAMPLE SEQUENCE IN EXP. I

Exposure	Function	Exposure	Function
HILL 7	dummy	ART 3	dummy
BOAT 9	presentation	YARD	test (6 sec.)
CAR 6	dummy	CAR 6	dummy
YARD 4	presentation	BOAT	test (16 sec.)
YARD 4	repetition	LOT 2	presentation
LINE 1	presentation	HILL 7	dummy
LINE	test (0 sec.)	LOT	test (2 sec.)

Results and Discussion

A single pairing of a word with a number resulted in proportions of recall of 1.00, .76, .40, and .24 after intervals of 0, 2, 6, and 16 sec., respectively. Immediate recall was not quite perfect since one error occurred. Recall decreased rapidly, but even at 16 sec. was well above the .11 chance level. A Friedman two-way analysis of variance resulted in a χ_r^2 of 40.9 which with 3 *df* is significant at the .001 level.

The proportion of recalls after 6 sec. for the two-presentation condition was .55. Seventeen *S*s recalled more after two presentations than after one, while only 3 *S*s recalled more after one presentation. Hence the difference in recall resulting from differing numbers of presentations was significant at the .01 level by two-tailed sign test.

The initial experiment indicates that retention of an association decreases over intervals less than those occurring between trials in list learning. It rarely happened that an association was not retained for any time at all. Any test of the number of associations produced by a single pairing would be influenced by the interval between presentation and test. In Exp. II a successive test procedure was utilized with varying intervals between Presentation, Test 1, and Test 2.

EXPERIMENT II

Method

The techniques of Exp. I were used with minor modifications. A memory drum taking a 32-in. tape was introduced, eliminating the necessity for changing the tape during an experimental session. Table 2 shows a sample block of exposures, the stimuli being trigrams

TABLE 2

SAMPLE SEQUENCE IN EXP. II

Exposure	Function	Exposure	Function
DAY 5	dummy	MAY 4	dummy
TAG 1	presentation	TAG	second test
BID 8	presentation	HAY 12	presentation
MAY 4	dummy	CAN 2	dummy
TAG	first test	BID	second test
DAY 5	dummy	ROT 7	presentation
CAN 2	dummy	ROT	first test
BID	first test		

of 99–100% Archer (1960) association value, and the responses being the numbers 1–15. Blocks varied in length from 31 to 35 exposures. A trigram and a number were presented together initially and tested by presenting the trigram alone on two later exposures. Interval A from presentation to Test 1 was either 0, 4, or 8 sec., while Interval B from Test 1 to Test 2 was either 4 or 8 sec. Each of the six combinations of these factors was tested 12 times with each of 36 introductory students. As in Exp. I, the presentation and tests of one condition overlapped that of other conditions. Three tapes assigning different random pairings of numbers and trigrams to the various conditions were made and each was used with 12 *Ss*.

Results

The mean proportions of numbers recalled on Test 1 were .99, .37, and .30 for the 0-, 4-, and 8-sec. conditions, respectively. Recall after a single pairing decreased more rapidly than in Exp. I, perhaps as a function of the increased number of alternative responses in Exp. II, where the chance level dropped to .07.

The analysis of recall on Test 2 is presented in Table 3. In all conditions, probabilities were calculated for individual *Ss* and means computed from these scores. The proportions recalled were inversely related to the length of both Intervals A and B. An analysis of variance showed that both Intervals A and B produced *F*s significant at the .001 level, using their respective interactions with *Ss* as the error terms ($F=14.9$, $df=2/70$; $F=14.0$, $df=1/35$). The AB interaction was also significant at the .001 level, using the triple interaction mean square as the error term ($F = 12.3$, $df = 2/70$). Recall on Test 2 was thus a function of the total elapsed time since presentation, and the usual finding of most rapid forgetting early in the interval immediately after presentation was also found here.

The comparisons of greatest interest are among the conditional probabilities of recall on Test 2 given a particular outcome on Test 1. The probability of a correct response on Test 2 given a correct response on Test 1 increased as the length of Interval A increased and decreased as Interval B increased. An analysis of variance found both these factors to be significant at the .001 level

TABLE 3

RECALL ON TEST 2, EXP. II

Interval B	Proportion Recalled			Probability of Correct Given Correct on Test 1			Probability of Correct Given Incorrect on Test 1		Probability of an Error Given that Error on Test 1	
	Interval A			Interval A			Interval A		Interval A	
	0 Sec.	4 Sec.	8 Sec.	0 Sec.	4 Sec.	8 Sec.	4 Sec.	8 Sec.	4 Sec.	8 Sec.
4 Sec.	.48	.34	.24	.49	.66	.71	.14	.06	.42	.49
8 Sec.	.30	.27	.25	.30	.46	.52	.29	.10	.28	.28

($F = 15.34$, $df = 2/70$; $F = 30.6$, $df = 1/35$), while their interaction F was less than 1. The interpretation of the effect of Interval B on cases where S is correct on Test 1 seems straightforward. In this case the beginning of the retention interval may be taken to be the removal of the test trigram from the drum aperture. Then the probability of recall decreases as the interval lengthens, in accord with the previously established functional relationship between recall probability and time. The increase in the dependent probabilities as Interval A lengthens suggests a selection of items. Those associations strong enough to survive to Test 1 over the longer intervals are those most likely to endure to Test 2. An alternate explanation would be that assuming Test 1 is a second presentation for pairs correctly recalled, then distributed presentations may be retained better than massed presentations. Either interpretation would seem to imply a continuum of associative strength.

The conditional probabilities of a correct response on Test 2 given an incorrect response on Test 1 decreased as Interval A increased, while these probabilities increased as Interval B increased. The 0-sec. conditions of Interval A are omitted because of the scarcity of incorrect responses in these conditions. An analysis of variance of the remaining conditions showed that Interval A was significant at the .001 level ($F = 8.81$, $df = 2/70$), while Interval B was significant at the .05 level ($F = 4.58$, $df = 1/35$). The interaction F was less than 1. A decrease in conditional probability as Interval A increases is understandable if Test 1 strengthens an interfering association. Test 1 would be expected to have greater relative influence the farther it is removed from the original presentation. The direct relationship between recall and the length of Interval B is also consistent with this interpretation. As time since pairing of cue and interfering response passes, the strength of this association decreases with the result that the relative strength of the original association increases. Thus a kind of spontaneous recovery of strength occurs that is in line with findings in retention experiments over longer intervals (Briggs, 1954). Strong evidence of interference from incorrect responses on Test 1 appeared in a comparison of the two conditions with equal time from presentation to Test 2. In the condition where Test 1 was closer to Test 2 than to presentation, the probability of a correct response on Test 2, given incorrect on Test 1, was markedly lower than in the other condition.

The strengthening effect of a test without knowledge of results may be inferred from the data on responses which were wrong on Test 1 and repeated on Test 2. Thirty-three *S*s made wrong responses in all four conditions. The conditional probabilities of the same error of commission being made on Test 2 as on Test 1 were computed for these *S*s and summarized in Table 3. A Friedman nonparametric two-way analysis of variance showed these to be significantly different at the .001 level ($\chi_r^2 = 16.6$, $df = 3$). The decrease in probability of recurrence as Interval B increased is suggestive of a retention effect following a contiguous pairing of stimulus and response. In this case S has determined the pairing instead of E, but the result seems similar to an E-arranged pairing. Note that the probability of an error recurring after 4 sec. was somewhat higher than the probability of recall of an E-paired response after

a corresponding interval, although the difference was not statistically significant. It is to be expected that all of the response alternatives are not equally likely to occur as errors, and therefore a selective factor may have influenced the occurrence of errors.

The data support the hypothesis that a test changes associative strength even though knowledge of results is withheld. When a response is correctly recalled, the test can be interpreted as a second pairing. Not only is the association reinstated at the zero point of the retention function, but the slope of the function is changed. When an error occurs at Test 1, a pairing of the stimulus with a new response sets the stage for retroactive interference. Thus, in successive tests, the probability of a right following a right would be expected to be high, while the probability of a right following a wrong should be low. This interpretation was tested further in Exp. III by having *E* either force a repetition of the original pairing or introduce a new pairing at a point in time equal to that of the first test in a double test condition.

Experiment III

Method

Four conditions were used with each of 28 *Ss* from the introductory psychology course. Condition DT (double test) reproduced the basic paradigm of Exp. II with a 4-sec. Interval A and a 4-sec. Interval B. Condition R (repetition) substituted a second presentation of the stimulus and response for Test 1 of the above condition. Condition I (interference) paired the original stimulus with a new response in place of Test 1 of Cond. DT. Condition S (single) tested retention once 10 sec. after a single presentation. Thus all four conditions were tested for recall 10 sec. after the original pairing. As in Exp. I and II, the conditions overlapped one another in each block.

The *S* was instructed to speak the number that had been paired with the word the first time that he saw it. Stimuli were common three- and four-letter words as in Exp. I. Responses were the numbers 1–12. Each block contained 29 exposures and 8 sec. elapsed between blocks. Each condition was tested twice within a block for eight blocks. Four tapes were prepared such that a particular pairing of word and number appeared equally often in each of the four conditions over the group of *Ss*.

Results and Discussion

The mean proportions recalled after 10 sec. in Cond. S and DT were .27 and .28, respectively, indicating that insertion of a test midway in the 10-sec. recall interval did not alter the overall proportion recalled. Any strengthening of associations correctly recalled is apparently offset by strengthening of erroneous responses. The mean proportion recalled in Cond. R was .56, showing that an *E*-controlled repetition did markedly increase recall. A two-tailed sign test was used to test this difference as well as the other comparisons in Exp. III. Condition R was superior to Cond. S for 25 of the 28 *Ss*, indicating a difference significant at the .01 level. The proportion recalled in Cond. I was .14, which was a marked decrease relative to Cond. S. This decrease after insertion of a different pairing was found with 21 *Ss*, while 3 *Ss* showed an increase, and 4 showed no change ($P < .01$).

The conditional probability in Cond. DT that a response was recalled on Test 2, given that it was correct on Test 1, was .68. The superiority of this to the mean proportion recalled in Cond. R was small, but it was found with 21 *Ss* ($P < .05$). The conditional probability that a response was recalled correctly on Test 2 given that it was incorrect on Test 1 was .10. This was only slightly lower than Cond. I, but the difference was found with 19 *Ss* and only 6 *Ss* showed the reverse difference ($P < .05$).

Thus the assumption that Test 1 strengthens or weakens the initially presented association depending on whether it is correctly or incorrectly recalled could account for the major part of the spreading of the conditional probabilities in Cond. DT but not all of it.

The presence of a selective factor is suggested by another comparison. The conditional probability in Cond. DT of an error on Test 2, given the same error on Test 1, was .46. This was appreciably higher than the proportion of second-presentation responses recalled in Cond. I, .27. Twenty *S*s showed the difference in this direction ($P < .05$). This suggests either differential initial probabilities of response to given items or differential difficulty of association, and it is only reasonable that a selective factor is also influencing correct responses on Test 1. Considering the joint evidence of Exp. II and III, the high probability of a correct recall on Test 2 given a correct recall on Test 1 appears to be due to the combined effect of selection and strengthening on Test 1. Individual differences among items are presumably related to past histories of *S*s. The effect of the test is to make the strong stronger and the weak weaker.

Experiment IV

The equivalence of a pairing read from the drum and one not so prompted was further investigated by testing short-term retention of an association after two pairings under several conditions of elicitation of the second pairing.

Method

Four conditions were used with each of 30 *S*s from the introductory course. Three conditions of repetition differed as to manner of production of the second pairing. The first pairing was produced by exposure of both stimulus and response in the aperture. In Cond. SR-P (SR-Prompt) both stimulus and response were shown again immediately after the first presentation. In Cond. S-P (S-Prompt) only the stimulus was shown on the exposure following the first presentation. In Cond. NP (No Prompt) neither stimulus nor response was shown *S*, but a set of ditto marks appeared following the first presentation. The *S* was instructed to read everything that appeared in the drum aperture, to supply the missing response whenever the stimulus appeared alone, and to repeat both stimulus and response just seen when the ditto marks appeared. An example of each type of presentation was shown *S*s on cards during the instructional preparation.

Recall was tested in each of these conditions by presenting the stimulus alone 6 sec. after the second presentation. In Cond. S (Single) recall was tested 6 sec. after one presentation.

The stimuli were trigrams of 99–100% Archer value. The responses were the numbers 1–10. Blocks were 26 exposures long, and 8 sec. elapsed between blocks. The three conditions of repetition were tested twice per block over 18 blocks, while Cond. S was tested once per block over 18 blocks. Conditions overlapped one another in each block. The order in which items were tested was randomized within each block. Three randomized assignments of pairs to conditions were made and each was used with 10 *S*s.

Results and Discussion

The mean proportions recalled were .42 ($\sigma_m = .02$) for Cond. SR-P, .44 ($\sigma_m = .02$) for Cond. S-P, .40 ($\sigma_m = .02$) for Cond. NP, and .35 ($\sigma_m = .03$) for Cond. S. The distributions of proportions for individual *S*s appeared to be approximately normal, and therefore a parametric test of significance was used. Conditions SR-P and S-P were both significantly different from Cond. S at the .01 level ($t = 3.1, 4.0$; $df = 29$), but did not differ significantly from each other. Condition NP was not significantly different from Cond. S, but was different from Cond. S-P at the .05 level ($t = 2.25$). The inferiority of Cond. NP was not due

to failure of *S*s to respond appropriately to the ditto marks, since no such failures were observed. The *S*'s behavior on second presentation was identical in all three conditions. Recall was well above the chance level of .10 for all conditions.

It may be concluded that *S*'s production of the response in the presence of the stimulus alone strengthens the association at least as much as reading the response from the drum. The difference in the two conditions in which the response was not prompted could perhaps be predicted from considerations of stimulus generalization. In Cond. S-P the stimulating situation is more similar to that obtaining in test for recall than is the case in Cond. NP.

The possibility of a progressive deterioration in performance over the experimental session (proactive interference) was investigated by summing correct responses in the three repetition conditions within successive combinations of three blocks. The proportions correct in successive stages were .42, .36, .44, .42, .43, and .46. A simple analysis of variance resulted in an F of 3.03 which was significant at the .05 level ($df = 5/145$). A t test between the first and last stages did not show a significant difference ($t = 1.34$, $df = 29$), and in any event the difference was in the wrong direction to be evidence for PI. There might be a transient interference effect in the second stage, but it is not marked. Progressive effects were further investigated by going back to the data of Exp. III and summing number correct after 10 sec. for all four conditions over successive pairs of blocks. The proportions obtained for the successive stages were .31, .28, .34, and .33. The F obtained in a simple analysis of variance was not significant, but again there is a slight dip in the second stage analyzed. Again, too, there is the possibility of a small practice effect from beginning to end of the session. In all the experiments described, any progressive effect was controlled by the equal appearance of all experimental conditions in each stage of the experiment.

Conclusions

It has been argued that short-term retention is qualitatively different from learning and long-term retention (Clark, Lansford, & Dallenbach, 1960). However, this study and an earlier one (Peterson & Peterson, 1959) have found some of the same phenomena in short-term retention that had previously been found in learning and retention studies over longer intervals, for example, increased recall through repetition, retroactive interference, and spontaneous recovery. It would seem parsimonious to assume that short-term retention is basically similar to retention over longer intervals until experimental evidence dictates otherwise.

The results of the four experiments taken together suggest that any interpretation of the successive test experiment must account for both the selection of items in relation to the passage of time and the effect of the first test in producing another pairing of the cue with some response. The change in the conditional probabilities on the second test as a function of the time between presentation and first test would seem to be a particularly recalcitrant characteristic for an all-or-none theory to accommodate. The results would appear to be most readily interpreted within the framework of an incremental theory of learning.

Summary

The formation of associations between pairs of verbal items was investigated in four short-term retention experiments. Experiment I showed retention to decrease regularly over a 16-sec. interval, although recall was nearly perfect immediately after presentation. In Exp. II the effect of a test without knowledge of results (Test 1) was found to vary with its position between original presentation (P)

and a second test (Test 2). Correct recall on Test 2, given correct recall on Test 1, increased as the time between P and Test 1 increased; correct recall on Test 2, given incorrect recall on Test 1, decreased as the time between P and Test 1 increased; error on Test 2, given that error on Test 1, decreased as the interval between tests increased. In Exp. III it was found that the strengthening or interference which resulted from substituting forced pairing of the same or different responses for Test 1 could not completely account for the diverging conditional probabilities in a successive test experiment. In Exp. IV the increase in recall from an unprompted response without knowledge of results was found to be as great as that for a prompted repetition.

The findings are discussed in relation to the all-or-none vs. incremental assumption regarding associative learning, and are interpreted as favoring the latter assumption.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74** (10, Whole No. 497).

BRIGGS, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47,** 285–293.

CLARK, L. L., LANSFORD, T. G., & DALLENBACH, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73,** 22–40.

ESTES, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67,** 207–223.

ESTES, W. K., HOPKINS, B. L., & CROTHERS, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60,** 329–339.

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *J. exp. Psychol.*, 1959, **58,** 193–198.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* New York: Lippincott, 1960.

(Received January 26, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 404–408

A TEST OF A GENERAL UTILITY THEORY MODEL FOR PROBABILITY LEARNING[1]

KENNETH B. LITTLE

University of Denver

YVONNE BRACKBILL AND STEPHEN H. KASSEL

Johns Hopkins University

The model presented in this report is an extension to the multichoice case of Siegel's (1959) Model I for probability learning in the two choice situation. The extended model asserts that in a noncontingent probability learning situation, the asymptotic response level of S will be a function of the relative frequency of occurrence, π_i, of the stimulus event within the set of stimulus events, of the marginal utility of making a correct response, a, and of the marginal utility for S of varying his responses, b.

In the general case of k alternatives, the expected utility of a correct prediction (U_r) is:

$$U_r = a \sum_{i=1}^{k} p_i \pi_i \qquad [1]$$

where k is the number of choices, and p_i is the proportion of correct responses for the event appearing with relative frequency π_i.

The expected utility of varying responses (U_v) is defined as b times the sum of the variances of p_i:

$$U_v = b \sum_{i=1}^{k} [p_i(1 - p_i)] \qquad [2]$$

This function has a maximal value when all $p_i = 1/k$, i.e., when S distributes his responses equally over all alternatives.

In order to find that value of p that maximizes total expected utility (U), it is necessary to solve the $k - 1$ simultaneous equations resulting from the $k - 1$ partial derivatives of the equation for U:

$$U = U_v + U_r = a \sum_{i=1}^{k} p_i \pi_i + b \sum_{i=1}^{k} [p_i(1 - p_i)] \qquad [3]$$

The solution is:

$$p_i = \frac{1}{k} + \frac{a}{b}\left[\frac{k\pi_i - 1}{2k}\right] \qquad [4]$$

When π_i and k are fixed the model indicates that S's strategy will be determined by the ratio of the marginal utilities, a/b. The value of this ratio, α, can be estimated from any set of probability learning data by the equation:[2]

$$\hat{\alpha} = 2\left[\frac{p_i - \frac{1}{k}}{\pi_i - \frac{1}{k}}\right] \qquad [5]$$

There is no a priori reason to assume that α will vary as a function of k or π. Consequently, once the α value for S is determined from one set of π and k values it should be possible to predict S's stable state response proportions for any other set of π and k values. As a test of

[1] This study was done while the second author held a United States Public Health Service Research Fellowship.

[2] This estimate differs from Siegel's (1959, p. 308) for the two-choice case by a factor of 2. The difference comes from using the sum of the variances rather than the average variance in the equation for U_v.

the model, the following experiment was conducted.[3]

Method

Subjects.—The *Ss* were 80 sixth grade students selected at random from classes within a single elementary school. The mean age was 12.0 yr. The *Ss* were assigned randomly to eight experimental groups, with the single restriction that equal numbers of boys and girls be included in each group. The groups represented different combinations of π and k values, the two variables of principal concern in this experiment. Also represented was a third variable of minor importance, even vs. uneven splits in the frequencies of occurrence of the less frequently occurring events. The eight conditions—together with the results—are presented in the upper part of Table 1.

Experimental material.—The stimuli were 11,200 3 × 5 in. colored cards. On one side of each card was printed a 1-in. high letter. The four letters used, E, T, A, and O, were chosen for their high frequency of occurrence in written English. Twenty-eight packs of 400 cards each were prepared for the eight conditions. For each pack of 400 cards, randomized series were prepared (and recorded) in blocks of 20 trials each. The particular objective probability split needed for any condition was maintained within each block (or approximately so, in the case of Treatments E and G). There were no other restrictions on randomization. Since the blocks were interchangeable, it was possible to make up a separately randomized order for each *S* in each group.

Procedure.—The *Ss* were tested individually in an experimental room in the school. The task was explained to *S*, and he understood that he was to guess the letter on the back of each card before it was turned over and to make as many correct predictions as possible. Trial 200 was followed by a rest period of approximately 2 min. Following Trial 400, *E* asked *S* to estimate the frequency of occurrence of each letter.

Results and Discussion

A preliminary analysis of variance for each of the eight groups indicated no significant differences among the means of the five blocks of trials beginning with Trial 301. By this criterion stable response levels had been reached. Table 1 shows, for each condition of the experiment, mean response proportions based on these final 100 trials.

In order to test the accuracy of prediction of the model, individual α scores were computed for each of the 80 *Ss* on the basis of their asymptotic response proportions to the most

TABLE 1

Mean Response Proportions for the Final 100 Trials

Study	Cond.	*N*	Measure	Choices			
				1	2	3	4
Basic	A	10	π	.500	.250	.250	
			p_o	.429	.299	.272	
	B	10	π	.500	.300	.200	
			p_o	.438	.302	.260	
	C	10	π	.600	.200	.200	
			p_o	.577	.253	.170	
	D	10	π	.600	.250	.150	
			p_o	.588	.211	.201	
	E	10	π	.500	.167	.167	.167
			p_o	.431	.178	.203	.188
	F	10	π	.500	.250	.150	.100
			p_o	.431	.235	.170	.164
	G	10	π	.600	.133	.133	.133
			p_o	.590	.134	.137	.139
	H	10	π	.600	.200	.150	.050
			p_o	.570	.156	.163	.111
Sup. A	E to C	8	π	.600	.200	.200	
			p_o	.615	.225	.160	
	B to G	8	π	.600	.133	.133	.133
			p_o	.561	.145	.148	.146
Sup. B	B	9	π	.500	.300	.200	
			p_o	.443	.271	.286	
	D	9	π	.600	.250	.150	
			p_o	.599	.231	.170	

[3] The authors are grateful for the cooperation of the Baltimore City Schools and for the help of J. T. Ahrens and Anthony Bravos in collecting the data for Supplementary Study B.

frequently appearing event, π_A. An initial analysis indicated that, as expected, there were no significant differences in the mean α values attributable to even vs. uneven splits in the π values of the less frequently occurring events. Therefore, the data were pooled for Groups A and B, C and D, E and F, and G and H. A two-way analysis of variance was then computed to determine the effects on α of major π value (.500 or .600) and number of choices ($k = 3$ or 4). The analysis indicated that α's based on the two levels of k were not significantly different ($F = .27$), but that there was a significant difference between α's as a function of π value ($F = 6.13, df = 1/76$, $P < .02$). Apparently, the model failed to predict accurately from one π value to the other.

This result was surprising in view of the relative accuracy of the model when applied to various data reported in the literature (Greenhouse, Little, Brackbill, & Kassel, 1960). Absolute differences between asymptotic response proportions obtained by Gardner (1957, 1958) and Cotton and Rechtschaffen (1958) and those predicted by the present model averaged only .029 even when k varied from 3 to 9 and π_A from .60 to .70. Since the major difference (with regard to the parameters of the model) between those studies and the present one was our use of a π_A value as low as .500, it seemed more than possible that the model's failure to predict was somehow associated with *Ss*' performance under the .500 π_A conditions. This suspicion was reinforced by the marked and consistent difference in closeness of probability matching between the four .500 π_A groups and the four .600 π_A groups. If the suspicion were true, it would mean that our original assumption was incorrect and that α is dependent on π.

However, there is an alternative explanation. In view of their age level, the possibility existed that *Ss* were unable to discriminate adequately among the frequencies of event occurrence in the .500 π_A series. Using the ratio of π_A to π_B (the next most frequently appearing event) as an index of discriminability, then on the average, the .500 π_A series presented more difficult tasks of discrimination than the .600 π_A series. However, if one examines the data in Table 1, it is clear that there is no systematic relationship between the magnitude of the p_A values and the π_A/π_B ratios, within either the .500 π_A or the .600 π_A sets. The *Ss* performed at about the same level when the ratios were small as when they were large. In addition, an examination of *Ss*' estimates of the relative frequency of occurrence of the stimuli revealed that their average error of estimate was somewhat larger for the .600 π_A series (.17) than for the .500 π_A series (.09).

Since both lines of evidence suggest that poor discrimination was not the factor responsible for the low response levels to π_A characterizing the .500 π_A series, we are left with the first proposition: that α is not independent of π. With low values of π_A, apparently the marginal utility of correct predictions is reduced, or the marginal utility of varying responses increases. A decrease in U_r is unlikely, in view of the results of a recent experiment by Brackbill, Kappy, and Starr (1962), who found a higher utility of correct predictions for an event of low frequency than one of high frequency.

It appears, then, that the model's failure to predict across π values is attributable to the fact that α *is* a

function of π under certain conditions. Specifically, the use of low values of π_A causes a disproportionate increase in the utility of varying responses; this, in turn, lowers the value of α, with the net result that predicted response levels are in error when one moves to higher π_A series.

Two supplementary studies were conducted to test the stability of the present results and the degree of confidence that could be placed in our interpretation of them.

Supplementary Study A.—The purpose of this study was to determine the extent of individual consistency in the utility of varying responses from one experimental condition to another, and also to ensure against the possibility that sampling error had in some unknown way contributed to the original results for the .500 π_A groups.

Eight each of the original *S*s tested under Cond. B and E were rerun under Cond. G and C, respectively, using the same experimental material and procedure as previously. Approximately 6 mo. intervened between testings.

The second part of Table 1 shows the mean response proportions obtained during the final 100 trials for the two groups. Individual α scores were computed from the obtained response proportions to the π_A event. These α scores were then compared, in a two-way analysis of variance, to the α scores of the 20 *S*s who were originally tested under Cond. C and G. (Number of choices, k, provided the second classification.) All values of F—groups, k, and interaction—were less than 1.00, clearly nonsignificant. These results reinforce our confidence in the adequacy of the original sampling procedure, and indirectly support the hypothesis that α is influenced by low values of π.

Despite a test-retest interval of 6 mo., the rank order correlation between the two sets of α scores for the 16 *S*s was .72 ($P < .01$). The ratio of the utility of correct responses to the utility of varying responses is apparently a fairly stable individual difference characteristic. In line with the argument presented earlier, this is more probably a reflection of individual differences in U_v than in U_r.

Supplementary Study B.—Two new groups of *S*s were tested on two occasions under Cond. B and D ($\pi = 50\text{-}30\text{-}20$ and $\pi = 60\text{-}25\text{-}15$). The purpose was twofold: (*a*) as a partial replication of the original experiment on the effects of major π value, and (*b*) to provide additional data on the stability of individual differences in response variability noted in Study A. Eighteen naive sixth grade *S*s were randomly assigned to two groups, which differed from each other only in terms of the sequence in which the two conditions were administered. The experimental material and procedure were the same as those used previously. A 1-wk. interval separated the two sessions for each *S*.

The lower part of Table 1 shows for both groups the mean response proportions obtained from the last 100 trials of the first testing session. Individual α scores were computed as before. A two-way analysis of variance was then used to compare α scores for the present group to the scores of the 20 *S*s run under Cond. B and D in the original experiment. Of the two variance classifications, experiments and π_A values, only the latter is significant ($F = 5.08, df = 1/32, P < .05$). As in Study A, there was a significant correlation between test and retest α scores ($\rho = .66, P < .01$).

Discussion

Considering the present experimental results together with those reported earlier (Greenhouse et al., 1960), the extension of Siegel's utility theory model probability learning must be considered only partly adequate. From α scores determined under one set of conditions, one can predict with reasonable accuracy the final response levels for other conditions that differ in the number of alternatives or choices. However, when the frequency of occurrence of the events is changed, prediction is more uncertain. The model is apparently satisfactory when $\pi_A = .600$ or .700 but fails when relative frequency of the major event drops to .500. This suggests that the model as originally proposed may hold only within a limited range of π_A values, and that the addition of a correction factor may be necessary for accurate prediction within a wider range.

It is interesting that *Ss* who received the π_A .500 series did not match probabilities in their asymptotic performance as closely as did *Ss* who received the π_A .600 series. Since multichoice noncontingent data have not been reported with π_A values of .500 for adults, it is unclear whether the phenomenon is a general one or reflects a developmental trend.

Summary

A utility theory model for probability learning in the multichoice, noncontingent case was derived as an extension of Siegel's (1959) Model I. To test the model, 80 sixth-grade children were tested individually in a probability learning experiment. The *Ss* were required to predict the occurrence of alphabetical letters on 400 consecutive trials. There were three experimental conditions: the relative frequency (π_A) of the most frequently appearing letter was either .500 or .600; the number of choices (k) was either 3 or 4; and the less frequently appearing letters were either evenly or unevenly divided in frequency of occurrence. Ten *Ss* were tested under each of the eight possible combinations of conditions thus formed.

The results indicated that (*a*) the model predicted asymptotic response proportions satisfactorily across number of choices but not across π values; (*b*) final response values for the most frequently appearing stimulus in the .500 π_A conditions were substantially below probability matching; and (*c*) the even or uneven division of frequency of occurrence of less frequently appearing stimuli did not significantly affect the response to the π_A letters. Subsidiary studies also demonstrated that α, defined as the ratio of the marginal utility of correct responses to the marginal utility of varying responses, is relatively stable over time for individual *Ss*, regardless of changes in π or k values.

References

Brackbill, Y., Kappy, M. S., & Starr, R. H. Magnitude of reward and probability learning. *J. exp. Psychol.*, 1962, **63**, 32–35.

Cotton, J. W., & Rechtschaffen, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

Gardner, R. A. Probability-learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

Gardner, R. A. Perception of relative frequency as a function of the number of stimulus and response categories. *USA Med. Res. Lab. Rep.*, 1958, No. 328.

Greenhouse, S., Little, K. B., Brackbill, Y., & Kassel, S. H. A model for multi-response choice probability learning. *Amer. Psychologist*, 1960, **15**, 456.

Siegel, S. Theoretical models of choice and strategy behavior: Stable state behavior in the two-choice uncertain outcome situation. *Psychometrika*, 1959, **24**, 303–316.

(Received February 6, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 409–412

PRESENTATION TIME, TOTAL TIME, AND MEDIATION IN PAIRED-ASSOCIATE LEARNING[1]

B. R. BUGELSKI

University of Buffalo

The present experiment deals with a neglected variable in the systematic study of memorization; namely, item presentation time per trial in relation to total learning time. This variable is obviously important in the interpretation of claims such as that of Rock (1957) for one-trial learning, claims that were based on an interitem time of 8 sec. For Rock, a trial consisted of 96 sec. during which 12 nonsense pairs were exposed. During this period *S* was free to repeat any item for the entire period, but even if he merely used the 8-sec. per item interval, he could still repeat any item quite frequently. The claim that repetition has no merit is obviously improperly drawn from such a study. Time per trial is also involved directly in studies of massed vs. spaced learning where massing is defined as a rapid interitem rate (Hovland, 1938). Using a rate of 2 sec. per item Hovland found learning inferior (14.89 trials) to a rate of 4 sec. per item (6.78 trials). A simple multiplication of trials by time, however, shows the two procedures to involve about the same total time (29.98 and 27.12 sec.). Even when a long intertrial interval is employed, total times appear to be about the same (22.36 and 23.40 sec.) for 2-sec. and 4-sec. presentation rates although the number of trials differed significantly (11.18 and 5.85).

[1] This study was supported by United States Public Health Research and Fellowship Branch Grant M-3489. The writer wishes to acknowledge the aid of N. Carriero and J. Rickwood in gathering the data.

It is the present hypothesis that in at least some areas of memorization, and under some conditions of presentation, the degree of learning will be a function of total time, regardless of the duration of the individual trials or interitem times. The hypothesis does not appear to be without support. Murdock (1960), employing free recall of words in lists of varying length, found total time to be a determining factor in how much was learned regardless of presentation rate. Attempts to compute total learning times from published studies have also proved suggestive for a variety of learning situations but the many variations in kinds and length of material and associated learning conditions prevent adequate comparisons. It is hoped that the present study will focus more attention on the various temporal factors in learning such as exposure time per item, interitem interval, intertrial interval, and total learning time. The essence of the problem is: what is a learning trial?

METHOD

Apparatus.—The Hunter Card Master was used to expose pairs of nonsense syllables at varying presentation times. The stimulus syllable was always exposed for 2 sec. and remained visible while the response syllable was exposed for either 2, 4, 6, 8, or 15 additional sec. Two seconds always elapsed before a new stimulus syllable appeared. Eight pairs of syllables which had been used by Rock, 1957 for his control group (see Table 2), were presented in three different random orders. The total time available to *S* for reacting to or learning a given pair, including the time before a new stimulus was presented,

varied, therefore, from 6 sec. for the 2-sec. response to 19 sec. for the 15-sec. response, with the other times lasting 8, 10, and 12 sec. per pair. The time allowed for anticipation of the response term was, of course, always 2 sec.

It will be noted that the total time for presentation of eight pairs of syllables, i.e., one trial, would last either 48, 64, 80, 96, or 152 sec. depending upon the exposure time of the response syllable. There was no rest period between trials.

Procedure.—Each *S* learned the same eight pairs of syllables to a criterion of two successful anticipations of the complete list. The apparatus used did not allow for the elimination of pairs as they were learned, but this was done in the computations so that the total learning times reported are based on the time taken to learn each pair to the criterion. Upon completion of learning each *S* was asked specifically how he happened to learn each pair, what made him think of the correct response, why some pairs were easy and some difficult.

Subjects.—One hundred sophomore students in experimental psychology courses, naive with respect to nonsense syllable learning, were assigned to one of the five experimental conditions, thereby making up five groups of 20. Assignment to groups depended only on order of arrival at the laboratory; the first *S* was assigned to Group 1, the next to Group 2, etc.

Results

Both trials and total time to learn are shown in Table 1 for each of the five different presentation times. The total times are obtained by multiplying the total number of trials to learn each pair in the series by the presentation time. As expected, the longer the presentation time per pair, the fewer the number of trials to criterion. Analysis of variance for the trials yields $F = 18.53$ ($P = .01$, $df = 99$). Total time, however, is not significantly different among the five conditions ($F = 0.845$). The correspondence between total times for the most rapid presentation time and the longest presentation time is striking. Although the former required more than three times as many trials as the latter, the mean time difference

TABLE 1

Mean Trials and Total Times to Learn Lists

Presentation Time	*N*	Trials		Total Exposure Time	
		Mean	*SD*	Mean	*SD*
6 sec.	20	10.2	4.1	61.2	24.3
8 sec.	20	8.8	3.8	70.1	30.4
10 sec.	20	5.8	1.9	57.9	19.0
12 sec.	20	4.7	2.5	56.1	29.6
19 sec.	20	3.3	1.2	62.2	22.2

is hardly more than 1 sec. Extrapolating roughly from these findings, it might be possible for similar *S*s to learn all eight pairs in one trial with a presentation time of about 25 sec. per pair. Within the limits, then, of the intervals and conditions used, the conclusion appears reasonable that presentation time multiplied by trials amounts to a constant value, $T_p \times T = k$.

During the postexperimental interrogation every *S* reported spontaneously that one kind of association, mediator or bridge occurred to him and was used to fix the responses. They did not, of course, use such terms and their explanations were ordered into a tentative and arbitrary classification of five seemingly discriminable groups such as the following:

I. The *S* formed one word of the two syllables, as, for example, from DUP-TEZ, he would create DEPUTIZE; from GAC-QET, he would form RACQUET or JACKET.

II. The syllables would be formed into two separate words or a phrase that could be initiated by the stimulus syllable: CEZ-MUN becomes SAYS MAN or C'EST MAN or SEND MONEY.[2]

III. The *S* would manufacture or imagine appropriate phrases or words that sound like the syllables, e.g., GEY-NUR becomes

[2] Categories I and II appear to be "sound bound." The associations appear to derive from the sounds of the syllables. The sounds are modified into words.

TABLE 2

MEAN TRIALS TO LEARN EACH PAIR OF ASSOCIATES AND THE CORRESPONDING NUMBER OF MEDIATORS VOLUNTEERED BY *Ss*, WITH BREAKDOWN BY TYPE OF MEDIATOR

Syllables	Trials		Total (out of a possible 100)	Mediators					
				Types					
	Mean	*SD*		I	II	III	IV	V	None
GEY-NUR	5.02	5.2	81	54	12	8	2	5	19
KAR-WEH	5.51	5.8	84	7	68	4	3	2	16
BIH-XIR	5.53	4.6	73	18	2	16	29	8	27
CEZ-MUN	6.79	6.2	70	6	31	13	4	16	30
FAX-SOQ	6.94	4.7	57	1	11	26	15	4	43
TOF-LAH	7.13	5.6	54	5	19	7	9	14	46
DUP-TEZ	7.16	5.8	52	2	1	32	5	12	48
GAC-QET	8.24	5.6	66	25	0	15	21	5	34

A GENERAL NURSE OR KAR-WEH becomes KAREN-WAYWARD. This category involves largely personal experience.

IV. Some *Ss* would attempt abstract analyses of the syllables to get at some lead to the response, e.g., BIH-XIR becomes BI = 2, *X* = 10, ONE NUMBER GOES WITH ANOTHER NUMBER, OR THEY BOTH HAD AN "I."

V. The *Ss* would report some vague association with one part of the combination, e.g., GAC-QET becomes translated into TOURNIQUET, BIH-XIR becomes SOMETHING LIKE BICEPS.

Not all pairs aroused reportable mediators for all *Ss* (see Table 2). Some mediators were forgotten after being used. Useless mediators were also commonly mentioned but those were impossible to classify and frequently forgotten. Many *Ss* reported inability to learn until a mediator of some type did occur. The longer the interitem interval the more likely the occurrence of a mediator. The five intervals in order from shortest to longest aroused respectively 94, 96, 109, 111, and 127 mediating associations.

DISCUSSION

The findings support the original hypothesis that total learning time is a significant variable to consider in at least some kinds of learning. While it may be convenient to break up a learning session into some kinds of units of time, and label these as "trials," it may be a questionable practice in trying to get at an appreciation of what *S* is actually learning or doing. If a presentation time is anything upwards of 1 sec., he may very well repeat the material to himself within the trial. To count the period as one trial may be quite inappropriate. To conclude, where *S* learns after one such interval, that he learned in one trial (implying no repetition) may be quite incorrect.

On the other hand, the reports from *Ss* support the interpretation that on many of the trials nothing effective is learned. Wrong responses may occur because inadequate mediators occurred to *S* or the mediators misled *S* into an inappropriate response. After varying amounts of search, which takes more or less time, an appropriate mediator might be generated or stimulated into functioning. Once this occurs, *S* might be able to respond correctly and meet the learning criterion. In a sense, then, learning does occur in one trial in such situations. It must be remembered that about a third of the time *S* is unable to report any mediator. On such occasions he reports "working on the pair" with a "it just came to me" type of answer. It may be that such unassignable asso-

ciations are of an unconscious nature, as suggested by Bugelski and Scharlock (1952), Russell and Storms (1955), and Ryan (1960).

The vast amount of research done on lists of pairs and on serial lists must be viewed from the limited viewpoint of how *lists* are learned, and not how pairs or other units are learned. From the viewpoint of a learning theory, it is the latter question that is of greater importance, however practical the former may be.

While the present study was not directed at the question of massed as opposed to spaced learning it does suggest that such studies too might take into consideration the question of total time involved in the learning sessions. Typically such studies provide some rest interval between trials for the spaced group while the massed *Ss* carry on with the activity. Commonly enough it is found that the spaced group performs at a higher level than the massed group after the same number of trials. When the spaced group's rests are included, however, it might be found that the total time was far in excess of the apparent advantage in trials, and the massed *Ss* have learned proportionally more than the spaced.

The present finding must be interpreted as restricted to the experimental conditions involved, particularly with respect to the time intervals employed, the method of exposure, and the nonsense material task. This report must be considered more suggestive than conclusive and other tasks must be examined before greater confidence can be placed in the constant time principle.[3]

[3] It has been suggested to the writer by the linguist, Henry Lee Smith, that there is a striking similarity in the ages at which children of different nationalities learn to speak their native language to about the same level of efficiency.

Summary

Five groups of 20 *Ss* learned eight pairs of nonsense syllables under conditions where the stimulus syllable was always presented for 2 sec. while the response syllable was presented for an additional 2, 4, 6, 8, or 15 sec. depending on the group. A 2-sec. interval preceded each new stimulus. The total exposure times for the several groups were 6, 8, 10, 12, or 19 sec. per syllable pair. A significant difference was found between trials to learn with the fastest learning occurring with the longest presentation times. When presentation time was multiplied by trials, however, no significant differences were found. It was concluded that T_p (presentation time) $\times$ T (trials) is a constant value.

Interrogation of *Ss* after learning revealed that in 67% of the total possible learning units, *Ss* made use of mediational devices. Those pairs that lent themselves most readily to translation into two meaningful words or could be combined into one word, or could initiate some imagery were most easily learned. These findings are discussed in terms of their significance for studies of "one-trial" learning.

References

Bugelski, B. R., & Scharlock, D. An experimental demonstration of unconscious mediation. *J. exp. Psychol.*, 1952, **44**, 334–338.

Hovland, C. I. Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **23**, 172–190.

Murdock, B. B. The immediate retention of unrelated words. *J. exp. Psychol.*, 1960, **60**, 222–234.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

Ryan, J. J. Comparison of verbal response transfer mediated by meaningfully similar and associated stimuli. *J. exp. Psychol.*, 1960, **60**, 408–415.

(Received February 21, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 413–414

SUPPLEMENTARY REPORT: GENERALIZATION OF A NONVERBAL RESPONSE TO AGGRESSIVE VERBAL STIMULI[1]

JAMES H. GEER AND ARNOLD H. BUSS
University of Pittsburgh

Buss (1961) investigated the effects of direction of generalization and intensity of vocal response on stimulus generalization with aggressive verbal stimuli. Generalizing from intense to mild aggressive stimuli yielded steeper gradients than generalizing from mild to intense stimuli. There was also a tendency to match the intensity of vocal response to the intensity of aggressive stimuli. The present study sought to test the generality of these findings with a nonverbal response and with an amplitude measure of response strength.

Method.—The response was a pull on the handle of an automobile parking brake mounted near *S*'s left knee at a 45° angle from the floor. The brake remained at its maximum excursion after each pull until deliberately returned to the starting point. The pause at the end of handle travel allowed *E* to read a pointer position on a scale mounted on the side of the apparatus, indicating the amount of handle travel. This scale value was converted to pounds of pull, the measure of response strength. Directly in front of *S* was a memory drum which presented the stimuli every 8 sec.

The *Ss* were pretrained to make "hard" and "soft" pulls on the brake. A hard pull was $\geq$30 lb. and a soft pull was $\leq$21 lb. All *Ss* were easily able to learn this discrimination and to make either response correctly upon request. The *Ss* were told to make only hard or soft pulls throughout the experiment.

Two groups of *Ss* generalized from intensely hostile stimuli, and their training series consisted of the 20 most hostile words. Group I was told "Right" for hard pulls and "Wrong" for soft pulls; Group III was told Right for soft pulls and Wrong for hard pulls. The other two groups generalized from mildly hostile stimuli, and their training series consisted of the 20 least hostile words. Group II was told Right for hard pulls and Wrong for soft pulls; Group IV was told Right for soft pulls and Wrong for hard pulls.

At the end of the 20 training trials *Ss* were instructed to continue responding, but *E* would no longer tell them whether they were right or wrong. Then the generalization series was begun. This series consisted of 126 words varying randomly from mild to intense hostility, as scaled by college *Ss* (Buss, 1961). The *Ss* were 60 undergraduate males enrolled in introductory psychology classes.

Results.—The *Ss* trained on high hostile words saw the word BLOODTHIRSTY first; the *Ss* trained on low hostile words saw the word DEBATE first. The amplitude of response to the first word prior to any reinforcement was categorized as "Strong" (>25 lb.) or "Weak" (<25 lb.) Twenty-four of the 30 *Ss* pulled strong to the word BLOODTHIRSTY and 21 of the 30 pulled weak to DEBATE. A chi square analysis of this distribution was significant beyond the .005 level. Thus there was a strong tendency to match response intensity with stimulus intensity prior to training.

Figure 1 shows the generalization gradients for the groups. The upper two gradients represent the hard pull groups; for these groups amplitude of response should diminish with distance from the training

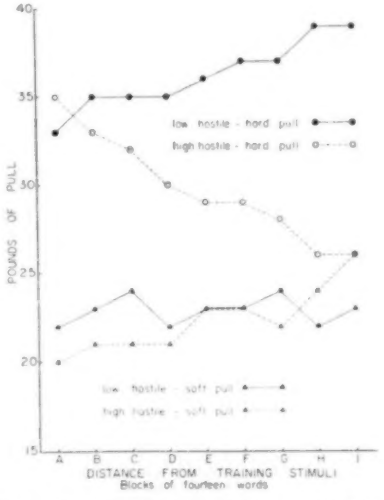

FIG. 1. Generalization gradients of response amplitude.

[1] Adapted from a MS thesis by the first author under the direction of the second author. The research was supported in part by Research Grant M-2681 from the National Institute of Mental Health, United States Public Health Service.

stimuli. The gradient of Group I does diminish, but Group II shows an inverse gradient: amplitude of response increases with distance from the training stimuli. An analysis of variance yielded a Groups × Stimuli F (the statistic relevant to slope differences) of 28.25 for 8 and 224 df ($P = .001$). Slopes were tested individually, and both were found to be significant ($P = .001$).

The lower two gradients represent the soft pull groups. Because the response measure is amplitude rather than frequency, the stimulus generalization gradients for these groups should rise, i.e., as the stimuli become more distant from the training stimuli the response should become less soft and therefore of greater amplitude. Group III's gradient does rise, but Group IV has a flat gradient. An analysis of variance yielded a Groups × Stimuli F of 4.21 for 8 and 224 df ($P = .001$). Slopes were tested individually; Group III's slope is significant, but Group IV's slope is not.

Discussion.—There was a tendency to match the intensity of the response with the intensity of the hostile stimuli. This was shown by both *S*s' responses to the first training word and the inverse gradient of Group II. Direction of generalization was also found to be a significant variable.

These two determinants, matching and direction of generalization, were also found to be important in Buss' (1961) study. He found that matching was a more potent determiner of the slope of generalization gradients than direction of generalization. In the present study direction of generalization was the more potent determiner: Groups I and III had steep gradients, whereas Groups II and IV had flat or inverse gradients. The reason for the discrepancy may be the nature of the response. It seems likely that the tendency to shout hostile words would be stronger than the tendency to pull harder on a brake handle to hostile words. Thus a verbal response, such as Buss used, would probably lead to closer matching of response intensity than a nonverbal response.

REFERENCE

BUSS, A. H. Stimulus generalization and aggressive verbal stimuli. *J. exp. Psychol.*, 1961, **61**, 469–473.

(Received March 31, 1961)

Journal of Experimental Psychology
1962, Vol. 63, No. 4, 414–415

SUPPLEMENTARY REPORT: STIMULUS GENERALIZATION GRADIENTS ALONG A LUMINOSITY CONTINUUM[1]

GARY OLSON AND RICHARD A. KING

University of North Carolina

Blough (1959), using pigeons, has obtained luminosity generalization gradients. He found relatively flat gradients with a tendency toward higher rates of responding at lower luminosities. The present experiment supplements Blough's findings.

Method.—The *S*s were 20 male Carneaux pigeons, approximately 100–120 days old when received from the supplier.

A Skinner box was used. The key, a ¾-in. diameter translucent disc, was illuminated by light produced by a slide projector with a 300-w. incandescent (ac) bulb. This light was focused by an additional convex lens and then passed through an Eastman Kodak optical wedge and balance wedge before striking the key. This optical wedge covered three log density units over the range of 15 setting positions. The settings used in this experiment were 2, 4, 6, 8 (CS), 10, 12, and 14 on the optical wedge. Each setting represents an increase of .4 log density units over the next smaller setting. Absolute luminosity values, obtained with a Macbeth illuminometer, were: Setting 2–13.89 mL.; Setting 8 (CS)-1.29 mL.; and Setting 14-0.11 mL. This is a much smaller range than that used by Blough. Blough also used a red colored disc.

The *S*s were key-peck trained after reaching 80% of their ad lib. weight. They were then put on a 1-min. variable interval (VI) schedule. Every minute of this training was divided into two periods: (*a*) a 50-sec. responding period with the key illuminated at Setting 8 (the CS); and (*b*) a 10-sec. blackout period in which there was no possibility of a reinforcement and the key was darkened by a shutter. Each daily session was 30 min. long.

[1] The data are taken from a thesis submitted to the Graduate School of the University of North Carolina in partial fulfillment of the requirements for the MA degree, Olson (1960).

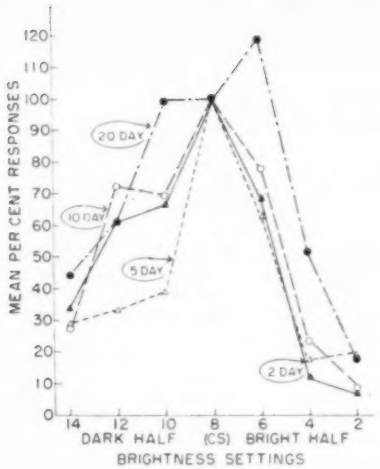

FIG. 1. Relative generalization gradients for the four training groups.

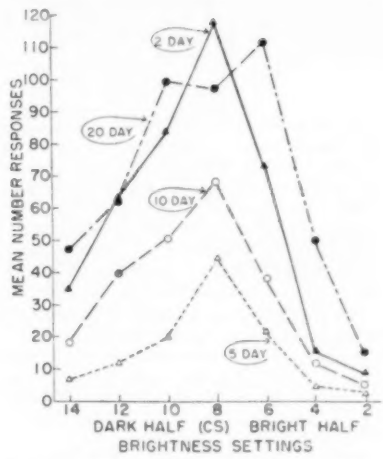

FIG. 2. Absolute generalization gradients for the four training groups.

The *S*s were taken directly from the lighted cage room to the darkened box, with no time allowed for dark adaptation. Since a second concern of this experiment was the effect of the amount of training on gradient slope, four groups of 5 *S*s each were given 20, 10, 5, or 2 days of VI training. The 2-day group consisted of *S*s specially selected for high response rate.

The generalization gradients were obtained during 2 consecutive days of extinction which began the day after the last training session. On Day 1 of extinction, there were 10 series of stimuli, and on Day 2, 5 series. Each stimulus series consisted of the seven light intensity values, ordered randomly. This is essentially the method of Guttman and Kalish (1956). The 50-sec. responding period, 10-sec. blackout cycle was maintained. During each 50-sec. period, *S* was given a different key illumination value. The wedge setting was changed during the 10-sec. blackout. The number of responses during each responding period was recorded, thus giving a total of 15 scores for each of the seven stimulus values.

Results.—The absolute and relative gradients for the four groups are shown in Fig. 1 and 2. The absolute measure is simply the total number of responses to each of the seven stimulus values. The relative measure is a percentage computed as follows:

$$\frac{\text{Total number of responses to a given stimulus value}}{\text{Total number of responses to the CS}} \times 100$$

The shapes of the gradients, like Blough's, are relatively flat. The statistical analyses indicated a significant dark preference for the absolute measure, but not the relative measure. This is similar to one of Blough's findings and is opposed to the stimulus intensity dynamism construct (Hull, 1952).

REFERENCES

BLOUGH, D. S. Generalization and preference on a stimulus-intensity continuum. *J. exp. Anal. Behav.*, 1959, **2**, 307–315.

GUTTMAN, N., & KALISH, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

OLSON, G. A study of two variables affecting stimulus generalization: Amount of training and stimulus intensity. Unpublished master's thesis, University of North Carolina, 1960.

(Received April 3, 1961)

Journal of Experimental Psychology
1962, Vol 63, No. 4, 416

SUPPLEMENTARY REPORT: SHIFT FROM NONREWARD TO NEGATIVELY CORRELATED REWARD[1]

FRANK A. LOGAN AND LOUIS M. GONZALEZ

Yale University

If naive rats are rewarded only if they take longer than a specified time running an alley, they gradually learn to adjust to this condition. Trapold (see Logan, 1960, p. 190) found that rats that had first run with reward uncorrelated with speed learned to run slowly as easily as naive rats. Subsequently, rats that had previously run without reward seemed to have unusual difficulty in adjusting to such a condition. The present study was undertaken to check this observation.

Method.—The *S*s were 24 Sprague-Dawley male rats, 120 days old at the beginning of the study. They were housed in individual cages and maintained on 12 gm/day Purina laboratory chow.

The double-alley maze and procedures used in this study have been described elsewhere (Logan, 1960). During Stage 1, each *S* in the experimental group was placed in a delay box, then run 1 or 4 trials in a 4-ft. straight alley without reward, and then placed in a second delay box before being returned to its home cage. This training lasted for 76 trials during which time *S*s in the control group were shuttled through the delay boxes each day but were not exposed to the alley. During Stage 2, all *S*s were run 6 massed trials a day and were rewarded with four 45-mg. Noyes pellets only if they took over 3 sec. in the alley, not counting starting time.

Results.—Performance by the experimental group during Stage 1 was independent of distribution of practice and is depicted by the X in Fig. 1. It should be noted that the rats took less time to run the alley without reward than they ultimately took when rewarded only for slow responses.

Performance curves of the two groups during Stage 2 are shown in Fig. 1. The experimental group initially ran somewhat faster than the control group, presumably because the exposure to the alley had reduced their fear and exploratory tendencies. However, the curves are closely comparable through most of their course and compare favorably with those obtained from naive *S*s. Hence, it is concluded that experience of nonreward does not affect adjustment to conditions of negatively correlated reward.

[1] Research supported in part by the National Science Foundation.

REFERENCE

LOGAN, F. A. *Incentive: How the conditions of reinforcement affect the performance of rats.* New Haven: Yale Univer. Press, 1960.

(Received April 22, 1961)

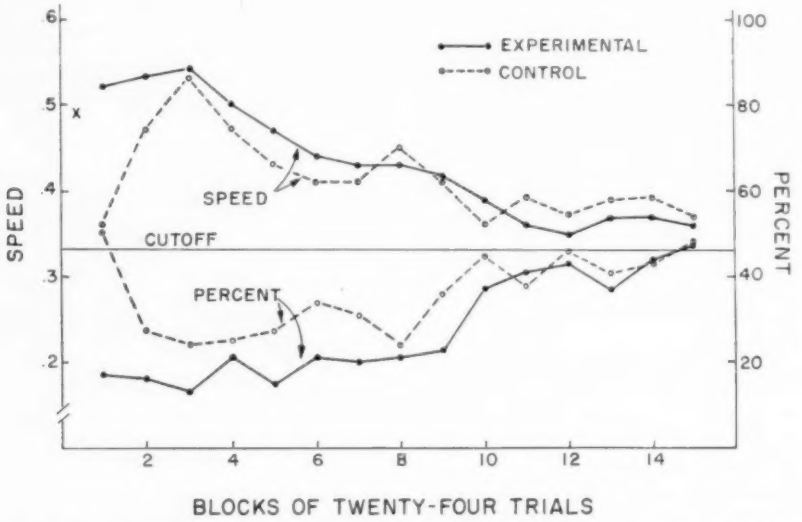

FIG. 1. Running speed and percentage below cutoff of *S*s that were rewarded only if they took over 3 sec. in the alley. (The experimental group had had prior exposure to the alley without reward, their performance at the end of this stage being indicated by the X; the control group had been handled but not previously exposed to the alley.)

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by BORIS SEMEONOFF

Vol. 53, Part 1 February, 1962 20s. net

B. B. JOYNSON and L. JOHN NEWSON. The perception of shape as a function of inclination.

C. G. COSTELLO. The effects of meprobamate on kinesthetic figural after-effects.

D. E. BERLYNE. Uncertainty and epistemic curiosity.

R. A. WILLETT and H. J. EYSENCK. Experimentally induced drive and difficulty level in serial rote learning.

GEORGE S. KLEIN, RILEY W. GARDNER and HERBERT J. SCHLESINGER: Tolerance for unrealistic experiences: a study of the generality of cognitive control.

R. B. CATTELL, J. HORN and H. J. BUTCHER: The dynamic structure of attitudes in adults: a description of some established factors and of their measurement by the motivational analysis test.

JOHN COHEN and E. J. DEARNALEY. Skill and judgment of footballers in attempting to score goals.

PUBLICATIONS RECENTLY RECEIVED:

The subscription price per volume, payable in advance, is 60s. net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS
Bentley House, Euston Road, London, N. W. 1

ZEITSCHRIFT FÜR
EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band VIII/4

HANS CH. DECHÊNE, **Über jugendlichen Reisedrang**

HARDI FISCHER UND CHARLES BUTSCH, **Musikalische Begabung und Intelligenz**

JOHANNES GRUNERT, **Unfallkeim—Unfalldisposition—Unfall**

WOLFGANG HORN, **Bedingungsfaktoren und Begleiterscheinungen wiederholter Straffälligkeit**

BERNHARD KRAAK, **Nicht-direktive Gruppentherapie mit Heimkindern**

WALTRAUT KÜPPERS, **Über eine Methode, Messungsergebnisse der psychogalvanischen Reaktion darzustellen**

HORST SACHER, **Der Moral-Charakter als Regulierungsgefüge der Persönlichkeit**

All articles are provided with summaries in English and French.
The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64,—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
(Brentanoweg 10, Germany)